GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: April 26, 2003, 13:02:26 ; Search time 79 Seconds
(without alignments)
1405.035 Million cell updates/sec

Title: US-10-027-000-2
Perfect score: 4391
Sequence: 1 MADIDVEAILKKLTLAEKVD.....DGVALRGKFTVGETYWWSGV 833

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	969	22.1	769	21	AAB18648	Amino acid sequenc
2	969	22.1	769	21	AAY67212	ORF 11 encoded par
3	953	21.7	721	19	AAW49862	Thermotoga maritim
4	951.5	21.7	3782	21	AAY77179	S. venezuelae deso
5	951.5	21.7	3782	23	AAE24228	Streptomyces venez
6	944.5	21.5	808	21	AAY77189	S. venezuelae macr
7	944.5	21.5	809	23	AAE24237	Streptomyces venez
8	907	20.7	721	18	AAW34558	Thermotoga maritim
9	869	19.8	752	17	AAR97199	Chimaeric thermost
10	860	19.6	735	19	AAW69761	Acetobacter xylinu
11	796.5	18.1	803	16	AAR85197	Tomatinase - a sap
12	747	17.0	778	19	AAW35004	Thermotoga maritim
13	731.5	16.7	744	13	AAR25384	Trichoderma reesei
14	731.5	16.7	744	21	AAB08340	Amino acid sequenc
15	731.5	16.7	744	21	AAY56548	Trichoderma reesei
16	728.5	16.6	793	16	AAR85199	Avenacinase-like p
17	698	15.9	793	16	AAR85200	Avenacinase-like p
18	681	15.5	792	16	AAR85198	Avenacinase - a sa
19	620	14.1	806	21	AAY77205	S. venezuelae deso
20	603.5	13.7	841	22	AAB82326	Aspergillus niger
21	603.5	13.7	860	22	AAB82320	Aspergillus niger
22	603.5	13.7	880	22	AAB82327	Aspergillus niger
23	603.5	13.7	883	22	AAB82328	Cell signal peptid
24	572.5	13.0	858	23	ABB07868	A. cellulolyticus
25	572	13.0	763	23	ABB90815	Herbicidally activ
26	564	12.8	756	23	ABB49932	Listeria monocytog
27	541.5	12.3	723	23	ABB48640	Listeria monocytog
28	532.5	12.1	774	23	ABB93754	Herbicidally activ
29	518.5	11.8	781	23	ABB92368	Herbicidally activ
30	512.5	11.7	868	21	AAG36535	Arabidopsis thalia
31	495.5	11.3	773	23	ABB93409	Herbicidally activ
32	495.5	11.3	1476	21	AAG49341	Arabidopsis thalia
33	491.5	11.2	792	23	ABB93421	Herbicidally activ
34	485	11.0	784	23	ABB93947	Herbicidally activ
35	457.5	10.4	739	21	AAG36536	Arabidopsis thalia
36	456.5	10.4	767	23	ABB91643	Herbicidally activ
37	455	10.4	728	21	AAG36537	Arabidopsis thalia
38	433.5	9.9	804	18	AAW10033	Beta-xylosidase (x
39	426	9.7	1328	21	AAG49342	Arabidopsis thalia
40	417.5	9.5	342	23	ABB90323	Human polypeptide
41	369.5	8.4	233	22	ABG30262	Novel human diagno
42	350.5	8.0	798	21	AAY52699	Aspergillus oryzae
43	333	7.6	650	23	ABB92769	Herbicidally activ
44	322.5	7.3	608	21	AAG54268	Arabidopsis thalia
45	322.5	7.3	608	23	ABB92570	Herbicidally activ

ALIGNMENTS

RESULT 1
AAB18648
ID AAB18648 standard; Protein; 769 AA.
XX
AC AAB18648;
XX
DT 22-JAN-2001 (first entry)
XX
DE Amino acid sequence of an ORF11 beta-glucosidase (desR).
XX
KW Narbonolide synthase; polyketide synthase gene; narbonolide polyketide;
KW antibiotic; C12-hydroxyalse; picK; desosamine biosynthesis;
KW desosaminyl transferase enzyme; ketolide; beta-glucosidase enzyme;
KW picromycin biosynthesis; beta-glucosidase.
XX
OS Streptomyces venezuelae.
XX
PN US6117659-A.
XX
PD 12-SEP-2000.
XX
PF 27-MAY-1999; 99US-0320878.
XX
PR 28-MAY-1998; 98US-0087080.
PR 22-SEP-1998; 98US-0100880.
PR 08-FEB-1999; 99US-0119139.
PR 20-MAY-1999; 99US-0134990.
PR 30-APR-1997; 97US-0846247.
PR 06-MAY-1998; 98US-0073538.
PR 28-AUG-1998; 98US-0141908.
XX
PA (KOSA-) KOSAN BIOSCIENCES INC.
XX

PI Ashley G, Betlach MC, Betlach M, Tang L, McDaniel R;
XX
DR WPI; 2000-610844/58.
XX
PT New recombinant picK hydroxylase gene of Streptomyces venezuelae useful
PT for converting ketolides to antibiotics and as antibiotics and
PT intermediates in the synthesis of compounds with pharmaceutical value
PT -
XX
PS Disclosure; Columns 37-38; 117pp; English.
XX
CC The present sequence represents a beta-glucosidase polypeptide.
CC The nucleotide sequence encoding it is used in the course of the
CC invention. The specification describes a recombinant DNA compound
CC expressing recombinant polyketide synthase genes in host cells for the
CC production of narbonolide, narbonolide derivatives and polyketides that
CC are useful as antibiotics and as intermediates in the synthesis of
CC compounds with pharmaceutical value. The DNA compounds may also encode
CC a C12-hydroxyalse (picK), desosamine biosynthesis and desosaminyl
CC transferase enzymes (useful for conversion of ketolides to antibiotics),
CC and the beta-glucosidase enzyme (involved in picromycin biosynthesis).
CC These compounds are also useful for increasing the antibiotic activity
CC of a compound relative to the unhydroxylated compound. The recombinant
CC host cells are useful as genetic systems that allow rapid engineering
CC of the narbonolide polyketide synthase. These would be valuable for
CC creating novel ketolide analogs for pharmaceutical applications.
XX
SQ Sequence 769 AA;

Query Match 22.1%; Score 969; DB 21; Length 769;
Best Local Similarity 32.6%; Pred. No. 1.5e-77;
Matches 277; Conservative 112; Mismatches 340; Indels 120; Gaps 25;

```
Qy     13 LTLAEKVDLLAGIDFW----------HTKALPKHGVPSLRFTDGPNGVRGTKFFNGVPAA 62
          :|| ||:  :     |          :  :|: |:| ||  |||||:|      |  |
Db      1 MTLDEKISFV----HWALDPDRQNVGYLPGVPRLGIPELRAADGPNGIR----LVGQTAT 52

Qy     63 CFPCGTSLGSTFNQTLLEEAGKMMGKEAIAKSAHVILGPTINMQRSPLGGRGFESIGEDP 122
            |   :| |||: |: :  ||:||::  | :  ::||| :|  | | ||| :|: |||
Db     53 ALPAPVALASTFDDTMADSYGKVMGRDGRALNQDMVLGPMMNNIRVPHGGRNYETFSEDP 112

Qy    123 FLAGLGAAALIRGIQSTGVQATIKHFLCNDQEDRRMMVQSIVTERALREIYALPFQIAVR 182
           ::   | | |:|||  |:  | |||  |:||: |  | : | |: |||    |: |
Db    113 LVSSRTAVAQIKGIQGAGLMTTAKHFAANNQENNRFSVNANVDEQTLREIEFPAFE-ASS 171

Qy    183 DSQPGAFMTAYNGINGVSCSENPKYLDGMLRKEWGWDGLIMSDWYGTYSTTEAVVAGLDL 242
           :  |:|| ||||:||     | : |: :|| :||: | :|||  |   |:|:  |||
Db    172 KAGAGSFMCAYNGLNGKPSCGNDELLNNVLRTQWGFQGWVMSDWLAT-PGTDAITKGLDQ 230

Qy    243 EM-----------PGPP-RFRGETLKFNVSNGK-PFIHVIDQRAREVLQFVKKCAASGV 288
          ||           | || :| || ||  | ||  |   |      | | |  |
Db    231 EMGVELPGDVPKGEPSPPAKFFGEALKTAVLNGTVPEAAVTRSAERIVGQMEKFGLLLAT 290

Qy    289 TENGPETTVNNTPETAALLRKVGNEGIVLLKNENNVLPLS--KKKKTLIVGPNAKQATYH 346
             ||    :     |: |||   | |||:||   |||:    |   ::|| |
Db    291 PAPRPE---RDKAGAQAVSRKVAENGAVLLRNEGQALPLAGDAGKSIAVIGPTAVDPKVT 347

Qy    347 GGGSAALRAYYAVTPFDGLSKQLETPPSYTVGAYTTVPPILGEQCLTPDGAPGMRWRVFN 406
          | ||| :    |  | | :  :          ||  ||    ||:          |    ||
Db    348 GLGSAHVVPDSAAAPLDTIKAR--------AGAGATVTYETGEETFGTQIPAGNLSPAFN 399

Qy    407 EPPGTPNRQHIDELFFTKTDMHLVDYYHPKAADTWYADMEGTYTADEDCTYELGLVVCGT 466
          :                  | ::   |  |   |   :|| |   |  | : :   |
Db    400 QG------------------HQLE---PGKAGALY---DGTLTVPADGEYRIAVRATG- 433

Qy    467 AKAYVDDQLVVDNATKQVPGDAFFGSATREETGRINLVKGNTYKFKIEFGSAPTYTLKGD 526
          |   ||    :  :     :|  :   :  : | || |:|           |: |
Db    434 --GYATVQL----GSHTIEAGQVYGKVS---SPLLKLTKG-THKL----------TISGF 473

Qy    527 TIVPGHGSLRVGGCKVIDDQAEIEKSVALAKEHDQVIICAGLNADWETEGADRASMKLPG 586
          :    || :|       | | |:|  |::    :: |    |  ||| || :: |||
Db    474 AMSATPLSLELGWVTPAAADATIAKAVESARKARTAVVFA---YDDGTEGVDRPNLSLPG 530
```

```
Qy    587 VLDQLIADVAAANPNTVVVMQTGTPEEMPWLDATPAVIQAWYGGNETGNSIADVVFGDYN 646
            |:||: || |||||:||: ||:   ||||  | ||:  || |     : | :::|| |
Db    531 TQDKLISAVADANPNTIVVLNTGSSVLMPWLSKTRAVLDMWYPGQAGAEATAALLYGDVN 590

Qy    647 PSGKLSLSFPKRLQDNPAFLNFRTEAG----------RTLYGEDVYVGYRYYEFADKDVN 696
          |||||: ||       ||  |    ||           :  | | ::|||||:::  :
Db    591 PSGKLTQSF-------PAAENQHAVAGDPTSYPGVDNQQTYREGIHVGYRWFDKENVKPL 643

Qy    697 FPFGHGLSYTTFAFSNLSVSH-KDGKLSVSLSVKNTGSVPGAQVAQLYVKPLQAAKINRP 755
          |||||||||||:|  |  :|     | | |:::|:|:|   | :| | |:          :
Db    644 FPFGHGLSYTSFTQSAPTVVRTSTGGLKVTVTVRNSGKRAGQEVVQAYLGASPNVTAPQA 703

Qy    756 VKELKGFAKVELQPGETKAVTIEEQEKYVAAYFDEERDQWCVEKGDYEVIVSDSSAAKDG 815
            |:| |: || |  || | ||:    : :  ::|   | |    |:  :    |||
Db    704 KKKLVGYTKVSLAAGEAKTVTVNVDRRQL-QFWDAATDNWKTGTGNRLLQTGSSSA---- 758

Qy    816 VALRGKFTV 824
            |||  ||
Db    759 -DLRGSATV 766
```

RESULT 2
AAY67212
ID AAY67212 standard; protein; 769 AA.
XX
AC AAY67212;
XX
DT 23-MAR-2000 (first entry)
XX
DE ORF 11 encoded partial beta-glucosidase, desR amino acid sequence.
XX
KW Narbonolide polyketide synthase; PKS; desosamine biosynthesis; desR;
KW beta-glucosidase; antibiotic production; narbomycin; picromycin;
KW ketolide.
XX
OS Streptomyces venezuelae.
XX
PN WO9961599-A2.
XX
PD 02-DEC-1999.
XX
PF 27-MAY-1999; 99WO-US11814.
XX
PR 28-MAY-1998; 98US-0087080.
PR 28-AUG-1998; 98US-0141908.
PR 22-SEP-1998; 98US-0100880.
PR 08-FEB-1999; 99US-0119139.
XX
PA (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI Ashley G, Betlach MC, Betlach M, McDaniel R, Tang L;
XX
DR WPI; 2000-072618/06.
DR N-PSDB; AAZ56002.
XX
PT New recombinant DNA encoding a domain of narbonolide polyketide
PT synthase, for production of ketolide antibiotics -
XX
PS Example 2; Page 33; 98pp; English.
XX
CC This is the Streptomyces venezuelae partial beta-glucosidase, desR amino
CC acid sequence, encoded by ORF11. This protein is involved in desosamine
CC biosynthesis. The invention relates to recombinant DNA containing a
CC coding sequence for a narbonolide polyketide synthase (PKS). Polyketides
CC are compounds synthesised from 2-carbon units through a series of
CC condensations and subsequent modifications. Modular PKSs are responsible
CC for the production of many antibiotics including picromycin. The
CC narbonolide PKS consists of a loading module, six extender modules, and
CC two thioester domains. Four proteins make up the narbonolide PKS (PICAI,
CC PICAII, PICAIII and PICAIV). PICAI includes the loading module and
CC extender modules 1 and 2, PICAII includes extender modules 3 and 4,
CC PICAIII includes extender module 5 and PICAIV includes extender module 6
CC and a type II thioesterase domain. The second type II thioesterase

```
CC   domain is found on the PICB protein. The nucleotide sequences encoding
CC   all of these proteins can be isolated in recombinant form from the
CC   recombinant cosmid pKOS023-27 (see AAZ56001). Narbonolide is
CC   desosaminylated in S. venezuelae to yield narbomycin, the desosaminyl
CC   transferase enzyme is required for this  conversion, and the desosamine
CC   biosynthetic genes are also found in cosmid pKOS023-27. The recombinant
CC   DNA of the invention is used to express, in transformed cells,
CC   narbonolide (or its derivatives) or other ketolides (particularly
CC   hybrids), which may then be converted (e.g. by other enzymes
CC   recombinantly expressed in the same hosts) to polyketide antibiotics or
CC   their intermediates. The antibiotics are useful in human or veterinary
CC   medicine.
XX
SQ   Sequence   769 AA;

  Query Match             22.1%;  Score 969;  DB 21;  Length 769;
  Best Local Similarity   32.6%;  Pred. No. 1.5e-77;
  Matches  277;  Conservative  112;  Mismatches  340;  Indels  120;  Gaps   25;

Qy      13 LTLAEKVDLLAGIDFW----------HTKALPKHGVPSLRFTDGPNGVRGTKFFNGVPAA 62
Db       1 MTLDEKISFV----HWALDPDRQNVGYLPGVPRLGIPELRAADGPNGIR----LVGQTAT 52

Qy      63 CFPCGTSLGSTFNQTLLEEAGKMMGKEAIAKSAHVILGPTINMQRSPLGGRGFESIGEDP 122
Db      53 ALPAPVALASTFDDTMADSYGKVMGRDGRALNQDMVLGPMMNNIRVPHGGRNYETFSEDP 112

Qy     123 FLAGLGAAALIRGIQSTGVQATIKHFLCNDQEDRRMMVQSIVTERALREIYALPFQIAVR 182
Db     113 LVSSRTAVAQIKGIQGAGLMTTAKHFAANNQENNRFSVNANVDEQTLREIEFPAFE-ASS 171

Qy     183 DSQPGAFMTAYNGINGVSCSENPKYLLDGMLRKEWGWDGLIMSDWYGTYSTTEAVVAGLDL 242
Db     172 KAGAGSFMCAYNGLNGKPSCGNDELLNNVLRTQWGFQGWVMSDWLAT-PGTDAITKGLDQ 230

Qy     243 EM----------PGPP-RFRGETLKFNVSNGK-PFIHVIDQRAREVLQFVKKCAASGV 288
Db     231 EMGVELPGDVPKGEPSPPAKFFGEALKTAVLNGTVPEAAVTRSAERIVGQMEKFGLLLAT 290

Qy     289 TENGPETTVNNTPETAALLRKVGNEGIVLLKNENNVLPLS--KKKKTLIVGPNAKQATYH 346
Db     291 PAPRPE---RDKAGAQAVSRKVAENGAVLLRNEGQALPLAGDAGKSIAVIGPTAVDPKVT 347

Qy     347 GGGSAALRAYYAVTPFDGLSKQLETPPSYTVGAYTTVPPILGEQCLTPDGAPGMRWRVFN 406
Db     348 GLGSAHVVPDSAAAPLDTIKAR--------AGAGATVTYETGEETFGTQIPAGNLSPAFN 399

Qy     407 EPPGTPNRQHIDELFFTKTDMHLVDYYHPKAADTWYADMEGTYTADEDCTYELGLVVCGT 466
Db     400 QG------------------HQLE---PGKAGALY---DGTLTVPADGEYRIAVRATG- 433

Qy     467 AKAYVDDQLVVDNATKQVPGDAFFGSATREETGRINLVKGNTYKFKIEFGSAPTYTLKGD 526
Db     434 --GYATVQL----GSHTIEAGQVYGKVS---SPLLKLTKG-THKL----------TISGF 473

Qy     527 TIVPGHGSLRVGGCKVIDDQAEIEKSVALAKEHDQVIICAGLNADWETEGADRASMKLPG 586
Db     474 AMSATPLSLELGWVTPAAADATIAKAVESARKARTAVVFA---YDDGTEGVDRPNLSLPG 530

Qy     587 VLDQLIADVAAANPNTVVVMQTGTPEEMPWLDATPAVIQAWYGGNETGNSIADVVFGDYN 646
Db     531 TQDKLISAVADANPNTIVVLNTGSSVLMPWLSKTRAVLDMWYPGQAGAEATAALLYGDVN 590

Qy     647 PSGKLSLSFPKRLQDNPAFLNFRTEAG----------RTLYGEDVYVGYRYYEFADKDVN 696
Db     591 PSGKLTQSF-------PAAENQHAVAGDPTSYPGVDNQQTYREGIHVGYRWFDKENVKPL 643

Qy     697 FPFGHGLSYTTFAFSNLSVSH-KDGKLSVSLSVKNTGSVPGAQVAQLYVKPLQAAKINRP 755
Db     644 FPFGHGLSYTSFTQSAPTVVRTSTGGLKVTVTVRNSGKRAGQEVVQAYLGASPNVTAPQA 703

Qy     756 VKELKGFAKVELQPGETKAVTIEEQEKYVAAYFDEERDQWCVEKGDYEVIVSDSSAAKDG 815
```

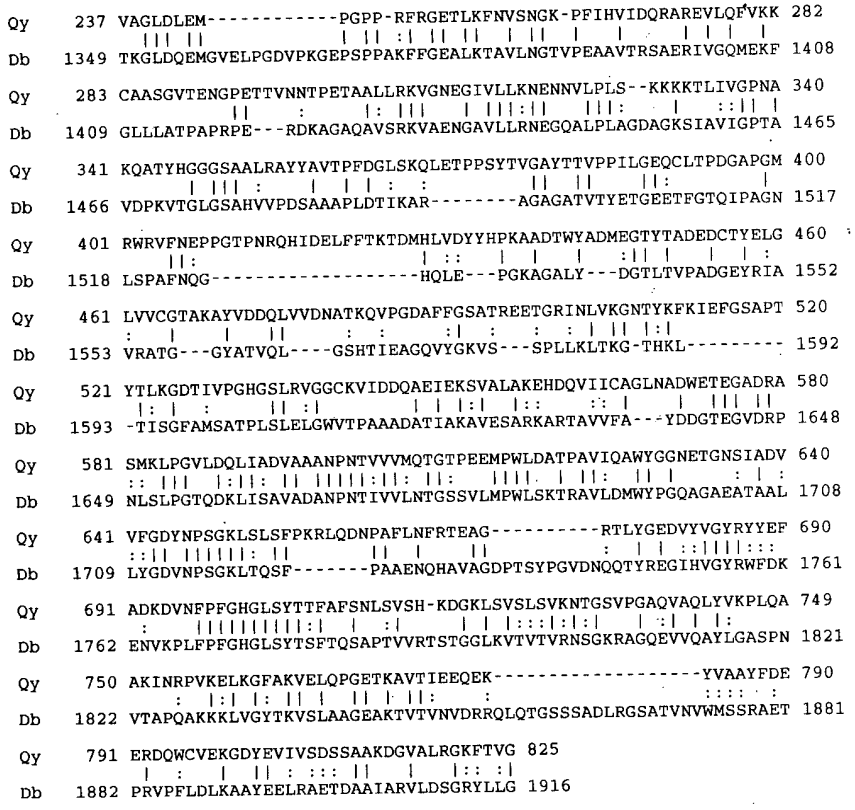
```
Db     704 KKKLVGYTKVSLAAGEAKTVTVNVDRRQL-QFWDAATDNWKTGTGNRLLQTGSSSA---- 758

Qy     816 VALRGKFTV 824
Db     759 -DLRGSATV 766


RESULT 3
AAW49862
ID   AAW49862 standard; Protein; 721 AA.
XX
AC   AAW49862;
XX
DT   21-DEC-1998  (first entry)
XX
DE   Thermotoga maritima MSB8-6G glycosidase.
XX
KW   Glycosidase; MSB8-6G; thermostable enzyme; oligosaccharide;
KW   glucose; sugar; baking; textile; detergent; beta-galactosidase.
XX
OS   Thermotoga maritima strain MSB8-6G.
XX
PN   WO9824799-A1.
XX
PD   11-JUN-1998.
XX
PF   08-DEC-1997;   97WO-US22623.
XX
PR   10-OCT-1997;   97US-0949026.
PR   06-DEC-1996;   96US-0056916.
XX
PA   (DIVE-) DIVERSA CORP.
XX
PI   Bylina EJ,  Lam DE,  Mathur EJ,  Swanson RV;
XX
DR   WPI; 1998-362407/31.
DR   N-PSDB; AAV36911.
XX
PT   Glycosidase enzymes from organisms of the genera Staphlothermus,
PT   Pyrococcus and Thermococcus  - for deriving sugar from
PT   oligosaccharides, useful in the e.g. food processing, textile or
PT   baking industries
XX
PS   Claim 1; Fig 5a-b; 92pp; English.
XX
CC   This is the amino acid sequence of glycosidase MSB6-6G, deduced
CC   from a polynucleotide (see AAV36911) of a Thermotoga maritima MSB8
CC   clone (6G) that grows optimally at 85 degC in high salt medium.
CC   The sequence shows 45% amino acid identity to beta-galactosidase
CC   bglB of Clostridium thermocellum.  The invention provides 18
CC   polynucleotides (see AAV36907-24) coding for thermostable glycosidases
CC   (see AAW49858-75) having glucosidase, alpha-galactosidase,
CC   beta-galactosidase, beta-mannosidase, beta-mannanase, endoglucanase
CC   or pullulanase activity.  Vectors and host cells are also claimed.
CC   A method is provided for producing the enzymes by recombinant
CC   techniques.  A claimed method for generating glucose from soluble
CC   cell oligosaccharides comprises contacting a sample (selected from
CC   dairy products, fruit juice, detergent, textile, guar gum, animal
CC   feed, plant biomass or waste product) containing oligosaccharides
CC   (selected from maltose, cellobiose, lactose, sucrose, raffinose,
CC   stachyose, verbascose, cellulose, starch, amylose, glycogen,
CC   disaccharides, polysaccharides and pullulan) with one of the
CC   claimed glycosidases such that glucose is produced.
XX
SQ   Sequence   721 AA;

  Query Match             21.7%;  Score 953;  DB 19;  Length 721;
  Best Local Similarity   30.3%;  Pred. No. 3.7e-76;
  Matches  261;  Conservative  127;  Mismatches  286;  Indels  186;  Gaps   25;

Qy       6 VEAILKKLTLAEKVDLLAGIDF----------WHTKALPKHGVPSLRFTDGPNGV 50
Db       4 IDEILSQLTTEEKVKLVVGVGLPGLFGNPHSRVAGAAGETHPVPRLGIPAFVLADGPAGL 63
```

```
Qy      51 R----GTKFFNGVPAACFPCGTSLGSTFNQTLLEEAGKMMGKEAIAKSAHVILGPTINMQ 106
           |         |       ||     | ||:|: |||| || ||:|        |:| | :|:
Db      64 RINPTRENDENTYYTTAFPVEIMLASTWNRDLLEEVGKAMGEEVREYGVDVLLAPAMNIH 123

Qy     107 RSPLGGRGFESIGEDPFLAGLGAAALIRGIQSTGVQATIKHFLCNDQEDRRMMVQSIVTE 166
           |:|| || ||   ||| |:|  |:| ::|:|| || | |||| |:|| |||  ||:| :||:|
Db     124 RNPLCGRNFEYYSEDPVLSGEMASAFVKGVQSQGVGACIKHFVANNQETNRMVVDTIVSE 183

Qy     167 RALREIYALPFQIAVRDSQPGAFMTAYNGINGVSCSENPKYLDGMLRKEWGWDGLIMSDW 226
           |||||||   |:|||: ::|   |:||| :|| ||:|   |  :||:|||: | :||||
Db     184 RALREIYLKGFEIAVKKARPWTVMSAYNKLNGKYCSQNEWLLKKVLREEWGFGGFVMSDW 243

Qy     227 YGTYSTTEAVVAGLDLEMPGPP-------RFRGETLKFNVSNGKPFIHVIDQRAREVLQF 279
           |   :  | : || |: |||          |   | :   :  ||    |:|:  | :|:
Db     244 YAGDNPVEQLKAGNDMIMPGKAYQVNTERRDEIEEIMEALKEGKLSEEVLDECVRNILK- 302

Qy     280 VKKCAASGVTENGPE---TTVNNTPET---AALLRKVGNEGIVLLKNENNVLPLSKKKKT 333
                 | | |       :| |:    | :  : | ||:|||:| | |||  :
Db     303 --------VLVNAPSFKGYRYSNKPDLESHAEVAYEAGAEGVVLLEN-NGVLPFDENTHV 353

Qy     334 LIVGPNAKQATYHGGGSAALRAYYAVTPFDGLSKQLETPPSYTVGAYTTVPPILGEQCLT 393
           : |   :    | ||      | ::  :|: :
Db     354 AVFGTGQIETIKGGTGSGDTHPRYTISILEGIKE------------------------- 387

Qy     394 PDGAPGMRWRVFNEPPGTPNRQHIDELFFTKTDMHLVDYYHPKAADTWYADMEGTYTADE 453
                 |   |:|   :   ::| :       |   : | |:  |:|
Db     388 -------RNMKFDEELASTYEEYIKK-------MRETEEYKPR-TDSW------------ 420

Qy     454 DCTYELGLVVCGTAKAYVDDQLVVDNATKQVPGDAFFGSATREETGRINLVKGNTYKFKI 513
               | |:    |  : :  : :   |: |                |: : |  |  |
Db     421 ------GTVI----KPKLPENFLSEKEIKKPPKKNDVAVVV------ISRISGEGYDRK- 463

Qy     514 EFGSAPTYTLKGDTIVPGHGSLRVGGCKVIDDQAEIEKSVALAKE-HDQVIICAGLNADW 572
              :|||  :                ||: |: |:|  :|| |||
Db     464 --------PVKGDFYLS------------DDELELIKTV--SKEFHDQ----------- 489

Qy     573 ETEGADRASMKLPGVLDQLIADVAAANPNTVVVMQTGTPEEM-PWLDATPAVIQAWYGGN 631
                                     ||::  |:| |:  | |    ::  |  |
Db     490 -------------------------GKKVVVLLNIGSPIEVASWRDLVDGILLVWQAGQ 523

Qy     632 ETGNSIADVVFGDYNPSGKLSLSFPKRLQDNPAFL---NFRTEAGRTLYGEDVYVGYRYY 688
           | |  :|||: |  ||||||  :|||   | |::      :    | :| ||:|||||||
Db     524 EMGRIVADVLVGKINPSGKLPTTFPKDYSDVPSWTFPGEPKDNPQRVVYEEDIYVGYRYY 583

Qy     689 EFADKDVNFPFGHGLSYTTFAFSNLSVSHKDGKLSVSLSVKNTGSVPGAQVAQLYVKPLQ 748
              :    :  : ||:|||| | : :| ::     | || :: |||   | :|:|:|:|
Db     584 DTFGVEPAYEFGYGLSYTKFEYKDLKIAIDGETLRVSYTITNTGDRAGKEVSQVYIK-AP 642

Qy     749 AAKINRPVKELKGFAKVE-LQPGETKAVTIEEQEKYVAAYFDEERDQWCVEKGDYEVIVS 807
             ||::| :||| | | : | |||:: :::|   : :|::  :|   |' || |:|||
Db     643 KGKIDKPFQELKAFHKTKLLNPGESEEISLEIPLRDLASFDGKE---WVVESGEYEVRVG 699

Qy     808 DSSAAKDGVALRGKFTV-GE 826
           ||     : ||  | | ||
Db     700 ASSR---DIRLRDIFLVEGE 716
```

RESULT 4
AAY77179
ID AAY77179 standard; Protein; 3782 AA.
XX
AC AAY77179;
XX
DT 05-JUN-2000 (first entry)
XX
DE S. venezuelae desosamine biosynthetic pikB-encoded protein, SEQ ID NO:4.
XX
KW Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW chronic obstructive pulmonary disease; respiratory inflammation;
KW hypercholesterolaemia; crop protection agent.
XX
OS Streptomyces venezuelae ATCC15439.
XX
PN WO200000620-A2.
XX
PD 06-JAN-2000.
XX
PF 25-JUN-1999; 99WO-US14398.
XX
PR 26-JUN-1998; 98US-0105537.
XX
PA (MINU) UNIV MINNESOTA.
XX
PI Sherman DH, Liu H, Xue Y, Zhao L;
XX
DR WPI; 2000-160679/14.
DR N-PSDB; AAZ87284.
XX
PT Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT synthesis of methymycin and pikromycin -
XX
PS Claim 19; Page 287-299; 438pp; English.
XX
CC The invention relates to an isolated and purified nucleic acid segment
CC comprising a desosamine biosynthetic gene cluster, a fragment or its
CC biologically active variant, where the nucleic acid sequence is not
CC derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC Streptomyces antibioticus. The invention also relates to a macrolide
CC biosynthetic gene cluster, or fragments thereof. The macrolide
CC biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC pikromycin, neomethymycin, narbomycin or a combination of these
CC compounds. Recombinant or augmented cells comprising the desosamine
CC and/or macrolide biosynthetic gene clusters are useful for the production
CC of biologically active macrolides. The macrolide biosynthetic proteins
CC are useful for synthesis of methymycin, pikromycin, neomethymycin and
CC narbomycin. The alternative termination of polyketide synthesis may be
CC useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
CC monomers. The compounds produced by the recombinant host cells are useful
CC as biopolymers, e.g., in packaging or biomedical applications, to
CC engineer PHA monomer synthases or to prepare biologically active agents,
CC such as chemotherapeutics, immunosuppressants, agents to treat asthma,
CC chronic obstructive pulmonary disease as well as other diseases involving
CC respiratory inflammation, cholesterol-lowering agents or macrolide-based
CC antibiotics which are active against a variety of organisms, e.g.,
CC bacteria, including multi-drug resistant pneumococci and other
CC respiratory pathogens, as well as viral parasitic pathogens, or as crop
CC protection agents (e.g., fungicides or insecticides) via expression of
CC polyketides in plants. The present sequence represents a protein
CC encoded by the desosamine biosynthetic gene cluster from Streptomyces
CC venezuelae ATCC 15439.
XX
SQ Sequence 3782 AA;

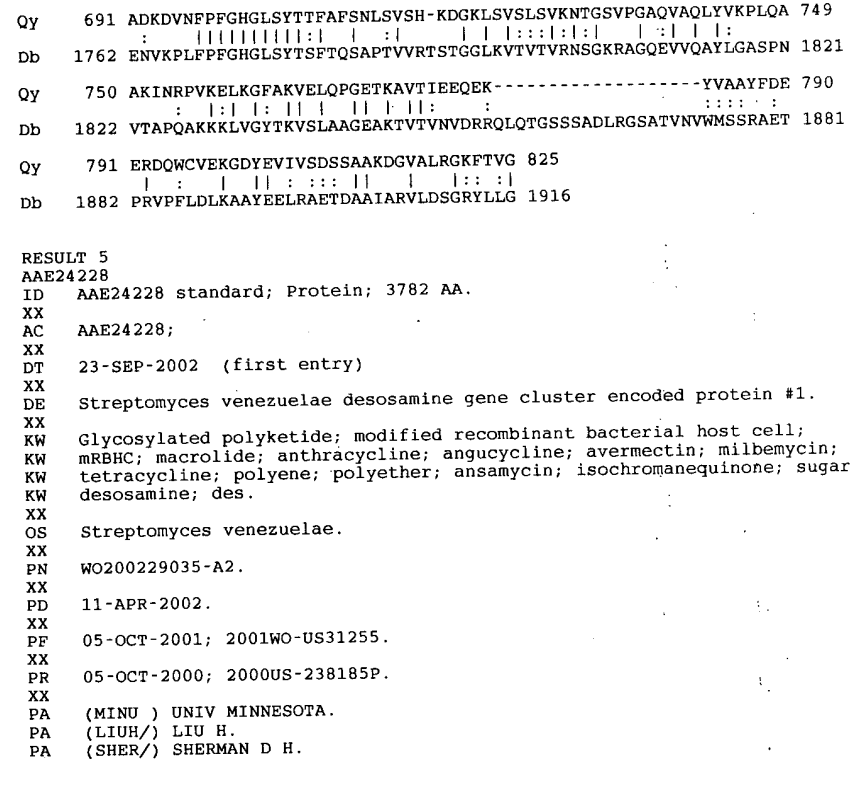
Query Match 21.7%; Score 951.5; DB 21; Length 3782;
Best Local Similarity 31.4%; Pred. No. 8.2e-75;
Matches 275; Conservative 123; Mismatches 340; Indels 137; Gaps 25;

```
Qy       9 ILKKLTLAEKVDLLAGIDFW----------HTKALPKHGVPSLRFTDGPNGVRGTKFFNG 58
             :: ::|| ||:  :     |          :   :|: |:| ||  |||||:|      |
Db    1121 LVAQMTLDEKISFV----HWALDPDRQNVGYLPGVPRLGIPELRAADGPNGIR----LVG 1172

Qy      59 VPAACFPCGTSLGSTFNQTLLEEAGKMMGKEAIAKSAHVILGPTINMQRSPLGGRGFESI 118
             |   |   :| |||: |: :  ||:||::  | :  ::||| :|  | | ||| :|:
Db    1173 QTATALPAPVALASTFDDTMADSYGKVMGRDGRALNQDMVLGPMMNNIRVPHGGRNYETF 1232

Qy     119 GEDPFLAGLGAAALIRGIQSTGVQATIKHFLCNDQEDRRMMVQSIVTERALREIYALPFQ 178
            ||| ::   | | |:||| |:  | ||| |:||: |  | : | |: |||    |:
Db    1233 SEDPLVSSRTAVAQIKGIQGAGLMTTAKHFAANNQENNRFSVNANVDEQTLREIEFPAFE 1292

Qy     179 IAVRDSQPGA--FMTAYNGINGVSCSENPKYLDGMLRKEWGWDGLIMSDWYGTYSTTEAV 236
           :   |: ||  || ||||:||     | : |: :|| :||: | :|||| |   |:|:
Db    1293 AS---SKAGAASFMCAYNGLNGKPSCGNDELLNNVLRTQWGFQGWVMSDWLAT-PGTDAI 1348
```



RESULT 5
AAE24228
ID AAE24228 standard; Protein; 3782 AA.
XX
AC AAE24228;
XX
DT 23-SEP-2002 (first entry)
XX
DE Streptomyces venezuelae desosamine gene cluster encoded protein #1.
XX
KW Glycosylated polyketide; modified recombinant bacterial host cell;
KW mRBHC; macrolide; anthracycline; angucycline; avermectin; milbemycin;
KW tetracycline; polyene; polyether; ansamycin; isochromanequinone; sugar;
KW desosamine; des.
XX
OS Streptomyces venezuelae.
XX
PN WO200229035-A2.
XX
PD 11-APR-2002.
XX
PF 05-OCT-2001; 2001WO-US31255.
XX
PR 05-OCT-2000; 2000US-238185P.
XX
PA (MINU) UNIV MINNESOTA.
PA (LIUH/) LIU H.
PA (SHER/) SHERMAN D H.
PA (ZHAO/) ZHAO L.
XX
PI Liu H, Sherman DH, Zhao L;
XX
DR WPI; 2002-405171/43.
XX
PT Modified recombinant bacterial host cells in which the expression and
PT activity of nucleic acids encoding sugar biosynthetic enzymes has been
PT altered, useful for producing metabolites with altered sugar structures
PT -
XX
PS Disclosure; Fig 8; 174pp; English.
XX
CC The invention provides a method to alter the sugar structure diversity
CC for a particular metabolite via the recruitment and collaborative action
CC of sugar genes from a variety of sugar biosynthetic pathways to yield a
CC metabolite comprising a non-natural sugar, e.g., a novel glycosylated
CC polyketide. The invention also relates to a modified recombinant
CC bacterial host cell (mRBHC) in which the expression and activity of
CC nucleic acids encoding sugar biosynthetic enzymes has been altered.
CC The mRBHCs may be cultured to produce the modified sugar products,
CC e.g. a macrolide, anthracycline, angucycline, avermectin, milbemycin,
CC tetracycline, polyene, polyether, ansamycin or isochromanequinone.
CC The present sequence is Streptomyces venezuelae sugar (desosamine)
CC biosynthetic gene cluster encoded protein.
XX
SQ Sequence 3782 AA;

Query Match 21.7%; Score 951.5; DB 23; Length 3782;
Best Local Similarity 31.4%; Pred. No. 8.2e-75;
Matches 275; Conservative 123; Mismatches 340; Indels 137; Gaps 25;



```
Db   1649 NLSLPGTQDKLISAVADANPNTIVVLNTGSSVLMPWLSKTRAVLDMWYPGQAGAEATAAL 1708

Qy    641 VFGDYNPSGKLSLSFPKRLQDNPAFLNFRTEAG----------RTLYGEDVYVGYRYYEF 690
Db   1709 LYGDVNPSGKLTQSF-------PAAENQHAVAGDPTSYPGVDNQQTYREGIHVGYRWFDK 1761

Qy    691 ADKDVNFPFGHGLSYTTFAFSNLSVSH-KDGKLSVSLSVKNTGSVPGAQVAQLYVKPLQA 749
Db   1762 ENVKPLFPFGHGLSYTSFTQSAPTVVRTSTGGLKVTVTVRNSGKRAGQEVVQAYLGASPN 1821

Qy    750 AKINRPVKELKGFAKVELQPGETKAVTIEEQEK------------------YVAAYFDE 790
Db   1822 VTAPQAKKKLVGYTKVSLAAGEAKTVTVNVDRRQLQTGSSSADLRGSATVNVWMSSRAET 1881

Qy    791 ERDQWCVEKGDYEVIVSDSSAAKDGVALRGKFTVG 825
Db   1882 PRVPFLDLKAAYEELRAETDAAIARVLDSGRYLLG 1916


RESULT 6
AAY77189
ID   AAY77189 standard; Protein; 808 AA.
XX
AC   AAY77189;
XX
DT   05-JUN-2000  (first entry)
XX
DE   S. venezuelae macrolide beta-glycosidase DesR, SEQ ID NO:24.
XX
KW   Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW   neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW   biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW   chronic obstructive pulmonary disease; respiratory inflammation;
KW   hypercholesterolaemia; crop protection agent.
XX
OS   Streptomyces venezuelae ATCC15439.
XX
PN   WO200000620-A2.
XX
PD   06-JAN-2000.
XX
PF   25-JUN-1999;   99WO-US14398.
XX
PR   26-JUN-1998;   98US-0105537.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Sherman DH,  Liu H,  Xue Y,  Zhao L;
XX
DR   WPI; 2000-160679/14.
DR   N-PSDB; AAZ87294.
XX
PT   Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT   synthesis of methymycin and pikromycin -
XX
PS   Claim 19; Page 371-373; 438pp; English.
XX
CC   The invention relates to an isolated and purified nucleic acid segment
CC   comprising a desosamine biosynthetic gene cluster, a fragment or its
CC   biologically active variant, where the nucleic acid sequence is not
CC   derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC   Streptomyces antibioticus. The invention also relates to a macrolide
CC   biosynthetic gene cluster, or fragments thereof. The macrolide
CC   biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC   pikromycin, neomethymycin, narbomycin or a combination of these
CC   compounds. Recombinant or augmented cells comprising the desosamine
CC   and/or macrolide biosynthetic gene clusters are useful for the
CC   production of biologically active macrolides. The macrolide biosynthetic
CC   proteins are useful for synthesis of methymycin, pikromycin,
CC   neomethymycin and narbomycin. The alternative termination of polyketide
CC   synthesis may be useful to prepare novel antibiotics and
CC   polyhydroxyalkanoate (PHA) monomers. The compounds produced by the
CC   recombinant host cells are useful as biopolymers, e.g., in packaging or
CC   biomedical applications, to engineer PHA monomer synthases or to prepare
CC   biologically active agents, such as chemotherapeutics,
CC   immunosuppressants, agents to treat asthma, chronic obstructive pulmonary
CC   disease as well as other diseases involving respiratory inflammation,
CC   cholesterol-lowering agents or macrolide-based antibiotics which are
CC   active against a variety of organisms, e.g., bacteria, including
CC   multi-drug resistant pneumococci and other respiratory pathogens, as well
CC   as viral parasitic pathogens, or as crop protection agents (e.g.,
CC   fungicides or insecticides) via expression of polyketides in plants.
CC   Sequences AAY77181-Y77189 represent desosamine biosynthetic enzymes from
CC   Streptomyces venezuelae ATCC 15439, which are encoded by sequences
CC   AAZ87286-Z87294.
XX
SQ   Sequence   808 AA;

  Query Match             21.5%;  Score 944.5;  DB 21;  Length 808;
  Best Local Similarity   32.2%;  Pred. No. 2.6e-75;
  Matches  275;  Conservative  112;  Mismatches  325;  Indels  143;  Gaps   26;

Qy      9 ILKKLTLAEKVDLLAGIDFW----------HTKALPKHGVPSLRFTDGPNGVRGTKFFNG 58
Db     55 LVAQMTLDEKISFV----HWALDPDRQNVGYLPGVPRLGIPELRAADGPNGIR----LVG 106

Qy     59 VPAACFPCGTSLGSTFNQTLLEEAGKMMGKEAIAKSAHVILGPTINMQRSPLGGRGFESI 118
Db    107 QTATALPAPVALASTFDDTMADSYGKVMGRDGRALNQDMVLGPMMNNIRVPHGGRNYETF 166

Qy    119 GEDPFLAGLGAAALIRGIQSTGVQATIKHFLCNDQEDRRMMVQSIVTERALREIYALPFQ 178
Db    167 SEDPLVSSRTAVAQIKGIQGAGLMTTAKHFAANNQENNRFSVNANVDEQTLREIEFPAFE 226

Qy    179 IAVRDSQPGA--FMTAYNGINGVSCSENPKYLDGMLRKEWGWDGLIMSDWYGTYSTTEAV 236
Db    227 AS---SKAGAASFMCAYNGLNGKPSCGNDELLNNVLRTQWGFQGWVMSDWLAT-PGTDAI 282

Qy    237 VAGLDLEM------------PGPP-RFRGETLKFNVSNGK-PFIHVIDQRAREVLQFVKK 282
Db    283 TKGLDQEMGVELPGDVPKGEPSPPAKFFGEALKTAVLNGTVPEAAVTRSAERIVGQMEKF 342

Qy    283 CAASGVTENGPETTVNNTPETAALLRKVGNEGIVLLKNENNVLPLS--KKKKTLIVGPNA 340
Db    343 GLLLATPAPRPE---RDKAGAQAVSRKVAENGAVLLRNEGQALPLAGDAGKSIAVIGPTA 399

Qy    341 KQATYHGGGSAALRAYYAVTPFDGLSKQLETPPSYTVGAYTTVPPILGEQCLTPDGAPGM 400
Db    400 VDPKVTGLGSAHVVPDSAAAPLDTIKAR--------AGAGATVTYETGEETFGTQIPAGN 451

Qy    401 RWRVFNEPPGTPNRQHIDELFFTKTDMHLVDYYHPKAADTWYADMEGTYTADEDCTYELG 460
Db    452 LSPAFNQG------------------HQLE---PGKAGALY---DGTLTVPADGEYRIA 486

Qy    461 LVVCGTAKAYVDDQLVVDNATKQVPGDAFFGSATREETGRINLVKGNTYKFKIEFGSAPT 520
Db    487 VRATG---GYATVQL----GSHTIEAGQVYGKVS---SPLLKLTKG-THKL--------- 526

Qy    521 YTLKGDTIVPGHGSLRVGGCKVIDDQAEIEKSVALAKEHDQVIICAGLNADWETEGADRA 580
Db    527 -TISGFAMSATPLSLELGWVTPAAADATIAKAVESARKARTAVVFA---YDDGTEGVDRP 582

Qy    581 SMKLPGVLDQLIADVAAANPNTVVVMQTGTPEEMPWLDATPAVIQAWYGGNETGNSIADV 640
Db    583 NLSLPGTQDKLISAVADANPNTIVVLNTGSSVLMPWLSKTRAVLDMWYPGQAGAEATAAL 642

Qy    641 VFGDYNPSGKLSLSFPKRLQDNPAFLNFRTEAG----------RTLYGEDVYVGYRYYEF 690
Db    643 LYGDVNPSGKLTQSF-------PAAENQHAVAGDPTSYPGVDNQQTYREGIHVGYRWFDK 695

Qy    691 ADKDVNFPFGHGLSYTTFAFSNLSVSH-KDGKLSVSLSVKNTGSVPGAQVAQLYVKPLQA 749
Db    696 ENVKPLFPFGHGLSYTSFTQSAPTVVRTSTGGLKVTVTVRNSGKRAGQEVVQAYLGASPN 755

Qy    750 AKINRPVKELKGFAKVELQPGETKAVTIEEQEKYVAAYFDEERDQWCVEKGDYEVIVSDS 809
```

```
                :  |:| |: || |  || | ||:                          |   : : |
Db     756 VTAPQAKKKLVGYTKVSLAAGEAKTVTVNV--------------------DRRQLQTGS 794

Qy     810 SAAKDGVALRGKFTV 824
           |:|     |||  ||
Db     795 SSAD----LRGSATV 805
```

RESULT 7
AAE24237
ID AAE24237 standard; Protein; 809 AA.
XX
AC AAE24237;
XX
DT 23-SEP-2002 (first entry)
XX
DE Streptomyces venezuelae DesR gene encoded protein.
XX
KW Glycosylated polyketide; modified recombinant bacterial host cell;
KW mRBHC; macrolide; anthracycline; angucycline; avermectin; milbemycin;
KW tetracycline; polyene; polyether; ansamycin; isochromanequinone; sugar;
KW desosamine; DesR; glucosidase; enzyme.
XX
OS Streptomyces venezuelae.
XX
FH Key Location/Qualifiers
FT Misc-difference 1
FT /note= "Encoded by GTG"
XX
PN WO200229035-A2.
XX
PD 11-APR-2002.
XX
PF 05-OCT-2001; 2001WO-US31255.
XX
PR 05-OCT-2000; 2000US-238185P.
XX
PA (MINU) UNIV MINNESOTA.
PA (LIUH/) LIU H.
PA (SHER/) SHERMAN D H.
PA (ZHAO/) ZHAO L.
XX
PI Liu H, Sherman DH, Zhao L;
XX
DR WPI; 2002-405171/43.
DR N-PSDB; AAD39043, AAD39052.
XX
PT Modified recombinant bacterial host cells in which the expression and
PT activity of nucleic acids encoding sugar biosynthetic enzymes has been
PT altered, useful for producing metabolites with altered sugar structures
PT -
XX
PS Disclosure; Page 171-173; 174pp; English.
XX
CC The invention provides a method to alter the sugar structure diversity
CC for a particular metabolite via the recruitment and cbllaborative action
CC of sugar genes from a variety of sugar biosynthetic pathways to yield a
CC metabolite comprising a non-natural sugar, e.g., a novel glycosylated
CC polyketide. The invention also relates to a modified recombinant
CC bacterial host cell (mRBHC) in which the expression and activity of
CC nucleic acids encoding sugar biosynthetic enzymes has been altered.
CC The mRBHCs may be cultured to produce the modified sugar products,
CC e.g. a macrolide, anthracycline, angucycline, avermectin, milbemycin,
CC tetracycline, polyene, polyether, ansamycin or isochromanequinone.
CC The present sequence is Streptomyces venezuelae sugar (desosamine)
CC biosynthetic gene cluster DesR (glucosidase) gene encoded protein.
XX
SQ Sequence 809 AA;

Query Match 21.5%; Score 944.5; DB 23; Length 809;
Best Local Similarity 32.2%; Pred. No. 2.6e-75;
Matches 275; Conservative 112; Mismatches 325; Indels 143; Gaps 26;

```
Qy       9 ILKKLTLAEKVDLLAGIDFW----------HTKALPKHGVPSLRFTDGPNGVRGTKFFNG 58
           :: ::|| ||:  :      |          :  :|: |:| ||  |||||:|      |
Db      56 LVAQMTLDEKISFV----HWALDPDRQNVGYLPGVPRLGIPELRAADGPNGIR----LVG 107

Qy      59 VPAACFPCGTSLGSTFNQTLLEEAGKMMGKEAIAKSAHVILGPTINMQRSPLGGRGFESI 118
             |   :| ||| |: :  ||:||::  | :  ::||| :|  | | ||| :|:
Db     108 QTATALPAPVALASTFDDTMADSYGKVMGRDGRALNQDMVLGPMMNNIRVPHGGRNYETF 167

Qy     119 GEDPFLAGLGAAALIRGIQSTGVQATIKHFLCNDQEDRRMMVQSIVTERALREIYALPFQ 178
           ||| ::  |.| |:||| |:  | ||| |:||: |  | : | |: |||| |:
Db     168 SEDPLVSSRTAVAQIKGIQGAGLMTTAKHFAANNQENNRFSVNANVDEQTLREIEFPAFE 227

Qy     179 IAVRDSQPGA--FMTAYNGINGVSCSENPKYLDGMLRKEWGWDGLIMSDWYGTYSTTEAV 236
             :  |: ||  || ||||:||     | : |: :|| :||: | :|||| |   |:|:
Db     228 AS---SKAGAASFMCAYNGLNGKPSCGNDELLNNVLRTQWGFQGWVMSDWLAT-PGTDAI 283

Qy     237 VAGLDLEM-----------PGPP-RFRGETLKFNVSNGK-PFIHVIDQRAREVLQFVKK 282
           ||| ||            | || :| || ||  | ||  |  |     | | |  |
Db     284 TKGLDQEMGVELPGDVPKGEPSPPAKFFGEALKTAVLNGTVPEAAVTRSAERIVGQMEKF 343

Qy     283 CAASGVTENGPETTVNNTPETAALLRKVGNEGIVLLKNENNVLPLS--KKKKTLIVGPNA 340
                 ||    :     |: |||   | |||:||   |||:    |  ::|| |
Db     344 GLLLATPAPRPE---RDKAGAQAVSRKVAENGAVLLRNEGQALPLAGDAGKSIAVIGPTA 400

Qy     341 KQATYHGGGSAALRAYYAVTPFDGLSKQLETPPSYTVGAYTTVPPILGEQCLTPDGAPGM 400
             | ||| :    |  | | :  :          ||  ||    ||:       |
Db     401 VDPKVTGLGSAHVVPDSAAAPLDTIKAR--------AGAGATVTYETGEETFGTQIPAGN 452

Qy     401 RWRVFNEPPGTPNRQHIDELFFTKTDMHLVDYYHPKAADTWYADMEGTYTADEDCTYELG 460
             ||:                          | ::   |  |   |   :|| |   |  | :
Db     453 LSPAFNQG------------------HQLE---PGKAGALY---DGTLTVPADGEYRIA 487

Qy     461 LVVCGTAKAYVDDQLVVDNATKQVPGDAFFGSATREETGRINLVKGNTYKFKIEFGSAPT 520
           :   |    |   ||     :  :      :|  :   :  : | || |:|
Db     488 VRATG---GYATVQL----GSHTIEAGQVYGKVS---SPLLKLTKG-THKL--------- 527

Qy     521 YTLKGDTIVPGHGSLRVGGCKVIDDQAEIEKSVALAKEHDQVIICAGLNADWETEGADRA 580
             |: |  :     || :|       | | |:|  |::    :: |    |  ||| ||
Db     528 -TISGFAMSATPLSLELGWVTPAAADATIAKAVESARKARTAVVFA---YDDGTEGVDRP 583

Qy     581 SMKLPGVLDQLIADVAAANPNTVVVMQTGTPEEMPWLDATPAVIQAWYGGNETGNSIADV 640
           :: |||  |:||: || ||||||:||: ||:   |||| | ||:  || |     : | :
Db     584 NLSLPGTQDKLISAVADANPNTIVVLNTGSSVLMPWLSKTRAVLDMWYPGQAGAEATAAL 643

Qy     641 VFGDYNPSGKLSLSFPKRLQDNPAFLNFRTEAG----------RTLYGEDVYVGYRYYEF 690
           ::|| |||||: ||       ||  |    ||          :  | | ::||||:::
Db     644 LYGDVNPSGKLTQSF-------PAAENQHAVAGDPTSYPGVDNQQTYREGIHVGYRWFDK 696

Qy     691 ADKDVNFPFGHGLSYTTFAFSNLSVSH-KDGKLSVSLSVKNTGSVPGAQVAQLYVKPLQA 749
             :   ||||||||||:|  |  :|     | | |:::|:|:|   | :| | |:
Db     697 ENVKPLFPFGHGLSYTSFTQSAPTVVRTSTGGLKVTVTVRNSGKRAGQEVVQAYLGASPN 756

Qy     750 AKINRPVKELKGFAKVELQPGETKAVTIEEQEKYVAAYFDEERDQWCVEKGDYEVIVSDS 809
           :  |:| |: || |  || | ||:                          |   : : |
Db     757 VTAPQAKKKLVGYTKVSLAAGEAKTVTVNV--------------------DRRQLQTGS 795

Qy     810 SAAKDGVALRGKFTV 824
           |:|     |||  ||
Db     796 SSAD----LRGSATV 806
```

RESULT 8
AAW34558
ID AAW34558 standard; Protein; 721 AA.
XX
AC AAW34558;
XX
DT 12-MAR-1998 (first entry)
XX
DE Thermotoga maritima MSB8 glycosidase.
XX
KW Glycosidase; thermostable; textile; food processing; pharmaceutical;
KW detergent; baking; industry; Thermococcus; Staphylothermus;

```
KW   Pyrococcus; glucose; soluble oligosaccharide.
XX
OS   Thermotoga maritima.
XX
PN   WO9725417-A1.
XX
PD   17-JUL-1997.
XX
PF   10-JAN-1997;   97WO-US00092.
XX
PR   13-SEP-1996;   96US-0712612.
PR   11-JAN-1996;   96US-0583787.
XX
PA   (RECO-) RECOMBINANT BIOCATALYSIS INC.
XX
PI   Bylina EJ, Lam DE, Mathur EJ, Swanson RV;
XX
DR   WPI; 1997-372858/34.
DR   N-PSDB; AAT93682.
XX
PT   New thermostable glycosidase(s) - from Thermococcus, Staphylothermus
PT   and Pyrococcus, used in the textile, food processing,
PT   pharmaceutical, detergent and baking industries
XX
PS   Claim 4; Fig 5; 82pp; English.
XX
CC   The present sequence represents glycosidase isolated from Thermotoga
CC   maritima. The enzyme or its encoding nucleic acid sequence is used for
CC   generating glucose from soluble oligosaccharides. The enzyme can be
CC   used in the food processing, pharmaceutical, textile, detergent and
CC   baking industries. The enzyme is also used to treat lactose intolerance,
CC   as a diagnostic reporter molecule, in corn wet milling or in the fruit
CC   juice industry. The enzymes can be used to hydrolyse guar gum to remove
CC   non-reducing terminal mannose residues. The nucleic acids encoding the
CC   enzyme may be used to generate probes to identify similar sequences.
XX
SQ   Sequence   721 AA;

  Query Match             20.7%;  Score 907;  DB 18;  Length 721;
  Best Local Similarity   30.4%;  Pred. No. 5.1e-72;
  Matches  263;  Conservative  121;  Mismatches  286;  Indels  194;  Gaps   29;

Qy       6 VEAILKKLTLAEKVDLLAGIDF----------------WHTKALPKHGVPSLRFTDGPNGV 50
           ::  ||  :||  |||  |:  |:                |  :|:  |:|:    |||  |:
Db       4 IDEILSQLTTEEKVKLVVGVGLPGLFGNPHSRVAGAAGETHPVPRLGIPAFVLADGPAGL 63

Qy      51 R----GTKFFNGVPAACFPCGTSLGSTFNQTLLEEAGKMMGKEAIAKSAHVILGPTINMQ 106
           |          |       ||     | ||:|: |||| || ||:|       |:|  | :|:
Db      64 RINPTRENDENTYYTTAFPVEIMLASTWNRDLLEEVGKAMGEEVREYGVDVLLAPAMNIH 123

Qy     107 RSPLGGRGFESIGEDPFLAGLGAAALIRGIQSTGVQATIKHFLCNDQEDRRMMVQSIVTE 166
           |:|| || ||   ||| |:|  |:| ::|:|| || | |||| |:|| ||:| :||:|
Db     124 RNPLCGRNFEYYSEDPVLSGEMASAFVKGVQSQGVGACIKHFVANNQETNRMVVDTIVSE 183

Qy     167 RALREIYALPFQIAVRDSQPGAFMTAYNGINGVSCSENPKYLDGMLRKEWGWDGLIMSDW 226
           |||||||   |:|||: ::|   |:||| :||  ||:|   |  :||:|||: | :||||
Db     184 RALREIYLKGFEIAVKKARPWTVMSAYNKLNGKYCSQNEWLLKKVLREEWGFGGFVMSDW 243

Qy     227 YGTYSTTEAVVAGLDLEMPGPP-------RFRGETLKFNVSNGKPFIHVIDQRAREVLQF 279
           |    :  | : || |: |||         |   | :   :  ||    |:|:  | :|:
Db     244 YAGDNPVEQLKAGNDMIMPGKAYQVNTERRDEIEEIMEALKEGKLSEEVLDECVRNILK- 302

Qy     280 VKKCAASGVTENGPE---TTVNNTPET---AALLRKVGNEGIVLLKNENNVLPLSKKKKT 333
                 |  | |       :| |:    | :  : | ||:|||:| | |||  :
Db     303 --------VLVNAPSFKGYRYSNKPDLESHAEVAYEAGAEGVVLLEN-NGVLPFDENTHV 353

Qy     334 LIVGPNAKQATYHGGGSAALRAYYAVTPFDGLSKQLETPPSYTVGAYTTVPPILGEQCLT 393
           : |    :    | ||                          :| | ||:   |
Db     354 AVFGTGQIETIKGGTGSG-----------------DTHPRYTISIETI---------- 384

Qy     394 PDGAPGMRWRVFNEPPGTPNRQHIDELFFTKTDMHLVDYYHPKAADTWYADMEGTYTADE 453
             |  |              |  |  | |  :  |  |   : | |:  |:|
Db     385 -KGGTG--------SGDTHPRYTISEEYIKK--MRETEEYKPR-TDSW---------- 420

Qy     454 DCTYELGLVVCGTAKAYVDDQLVVDNATKQVPGDAFFGSATREETGRINLVKGNTYKFKI 513
                 | |:    |  : :  :  :  |: |                  |: : |  |  |
Db     421 ------GTVI----KPKLPENFLSEKEIKKPPKKNDVAVVV------ISRISGEGYDRK- 463

Qy     514 EFGSAPTYTLKGDTIVPGHGSLRVGGCKVIDDQAEIEKSVALAKE-HDQVIICASLNADW 572
                 :|||  :              ||: |: |:|  :|| |||
Db     464 --------PVKGDFYLS-----------DDELELIKTV--SKEFHDQ---------- 489

Qy     573 ETEGADRASMKLPGVLDQLIADVAAANPNTVVVMQTGTPEEM-PWLDATPAVIQAWYGGN 631
                                        ||::  |:| |:  | |    ::  |  |
Db     490 -------------------------GKKVVVLLNIGSPIEVASWRDLVDGILLVWQAGQ 523

Qy     632 ETGNSIADVVFGDYNPSGKLSLSFPKRLQDNPAFLNFRTEAGR--TLYGEDV-----YVG 684
           | |  :|||: |  ||||||  :|||   | |::    |  |:  | : :|      :
Db     524 EMGRIVADVLVGKINPSGKLPTTFPKDYSDVPSW----TFPGKLPTTFPKDYSDVPSWTF 579

Qy     685 YRYYEFADKDVNFPFGHGLSYTTFAFSNLSVSHKDGKLSVSLSVKNTGSVPGAQVAQLYV 744
           |||:    :  : ||:|||||  | : :| ::     | || :: |||   | :|:|:|:
Db     580 PRYYDTFGVEPAYEFGYGLSYTKFEYKDLKIAIDGETLRVSYTITNTGDRAGKEVSQVYI 639

Qy     745 KPLQAAKINRPVKELKGFAKVE-LQPGETKAVTIEEQEKYVAAYFDEERDQWCVEKGDYE 803
           |     ||:|  :||| | | : | |||:: :::|   : :|::  :|   | || |:||
Db     640 K-APKGKIDKPFQELKAFHKTKLLNPGESEEISLEIPLRDLASFDGKE---WVVESGEYE 695

Qy     804 VIVSDSSAAKDGVALRGKFTV-GE 826
           | |  ||     : ||  | | ||
Db     696 VRVGASSR---DIRLRDIFLVEGE 716


RESULT 9
AAR97199
ID   AAR97199 standard; Protein; 752 AA.
XX
AC   AAR97199;
XX
DT   18-OCT-1996  (first entry)
XX
DE   Chimaeric thermostable beta-glucosidase.
XX
KW   Chimaera; beta-glucosidase; Cellvibrio gilvus; Agrobacterium tumefaciens;
KW   homology; decomposed cellulose; glucose; cellobiose; cellotriose;
KW   cellotetraose; cellopentaose.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Region          1..692
FT                   /note= "amino acids derived from Cellvibrio gilvus
FT                        beta-glucosidase"
FT   Region          693..752
FT                   /note= "amino acids derived from Agrobacterium
FT                        tumefaciens beta-glucosidase residues 759-818"
XX
PN   JP08131168-A.
XX
PD   28-MAY-1996.
XX
PF   09-NOV-1994;   94JP-0299049.
XX
PR   09-NOV-1994;   94JP-0299049.
XX
PA   (NORQ ) NORINSUISANSHO SHOKUHIN SOGO.
XX
DR   WPI; 1996-311636/32.
DR   N-PSDB; AAT32999.
XX
PT   Thermostable chimaeric beta-glucosidase - useful for prodn. of
PT   cellobiose from decomposed cellulose
XX
PS   Claim 2; Page 5-8; 9pp; Japanese.
XX
CC   This is the amino acid sequence of a novel chimaeric beta-glucosidase
```

```
CC   enzyme composed of amino acids 1-692 of Cellvibrio gilvus
CC   beta-glucosidase and amino acids 759-818 from the Agrobacterium
CC   tumefaciens beta-glucosidase, replacing amino acids 693-752 of the
CC   Cellvibrio sequence.  The chimaera was constructed by isolating the
CC   C.gilvus beta-glucosidase gene and comparing its sequence with the
CC   corresp. genes from several other species, e.g. Hansenula anomala,
CC   Ruminococcus albus, Butyrivibrio fibrisovens, Agrobacterium tumefaciens,
CC   etc. The homologous sequences from the other species could then be used
CC   to generate chimaeric glucosidase genes.  The novel chimaeric protein
CC   acts on decomposed cellulose to generate glucose and cellobiose and has
CC   a reduced ability to decompose cellobiose as compared to cellotriose,
CC   cellotetraose or cellopentaose as substrate.  The novel enzyme has an
CC   optimum pH of 6.0 and temp of 40-45 deg. C, is stable in pH 4-9 and at
CC   35 deg. C.
XX
SQ   Sequence   752 AA;

  Query Match             19.8%;  Score 869;  DB 17;  Length 752;
  Best Local Similarity   29.2%;  Pred. No. 1.4e-68;
  Matches  250;  Conservative  107;  Mismatches  279;  Indels  220;  Gaps   22;

Qy       2 ADIDVEAILKKLTLAEKVDLLAGI---DFW----HTKALPKH----------GVPSLRFT 44
Db      51 ADQRARLAVQAMTQQEKLRWVFGYFGHDFGKSKKHPDALPQSAGYIPGTPRLGLPALFET 110

Qy      45 DGPNGVRGTKFFNGVPAACFPCGTSLGSTFNQTLLEEAGKMMGKEAIAKSAHVILGPTIN 104
Db     111 DAGQGVASQSGANVRERTALPSGLSTASTWDPKVAYAGGAMIGSEARASGFNVMLAGGVN 170

Qy     105 MQRSPLGGRGFESIGEDPFLAGLGAAALIRGIQSTGVQATIKHFLCNDQEDRRMMVQSIV 164
Db     171 LQREPRNGRNFEYAGEDPLLAGTMIGQAIKGVESNRIISTLKHFVLNDQETGRNELDARI 230

Qy     165 TERALREIYALPFQIAVRDSQPGAFMTAYNGINGVSCSENPKYLDGMLRKEWGWDGLIMS 224
Db     231 DKAALRMSDLLAMELALEQSDAGSVMCAYNRLNGPYTCEHPWLLSEVLKRDWGFRGYVMS 290

Qy     225 DWYGTYSTTEAVVAGLD----LEMPGPPRFRGETLKFNVSNGKPFIHVIDQRARE--VLQ 278
Db     291 DWGATHSTVAAANSGLDQQSGQEFDKSPYF-GGALEEAVKTG-----AVPQKRLDDMVTR 344

Qy     279 FVKKCAASGVTENGPETTV--------NNTPETAALLRKVGNEGIVLLKNENNVLPLSKK 330
Db     345 IVRTMFGKGVVDNPLKPGVAIDFAANGGQPPD--------GEEGMVLLKNEGRLLPLAKT 396

Qy     331 KKTL-IVGPNAKQATYHGGGSAALRAYYAVTPFDGLSKQLETPPSYTVGAYTTVPPILGE 389
Db     397 VRTIAVIGGHADAGVLSGGGSSQV--------------------YPVGG-------IAV 428

Qy     390 QCLTPDGAPGMRWRVFNEPPGTPNRQHIDELFFTKTDMHLVDYYHPKAADTWYADMEGTY 449
Db     429 KGLLPATWPG-----------------PVVYYPSSPLRAIQAQAPNA------------ 458

Qy     450 TADEDCTYELGLVVCGTAKAYVDDQLVVDNATKQVPGDAFFGSATREETGRINLVKGNTY 509
Db     459 ------------------------------------------------------------ 458

Qy     510 KFKIEFGSAPTYTLKGDTIVPGHGSLRVGGCKVIDDQAEIEKSVALAKEHDQVIICAGLN 569
Db     459 --KV--------------------------VFDDGRDPARAARVAAGADVALVFAN-- 486

Qy     570 ADWETEGADRASMKLPGVLDQLIADVAAANPNTVVVMQTGTPEEMPWLDATPAVIQAWYG 629
Db     487 -QWIGEANDAQTLALPDGQEELITSVAGANGRTVVVLQTGGPVTMPWLARVPAVLEAWYP 545

Qy     630 GNETGNSIADVVFGDYNPSGKLSLSFPKRLQD-----------NPAFLNFRTEAGRTLYG 678
Db     546 GTSGGEAIANVLFGAVNPSGHLPATFPQSEQQLPRPKLDGDPKNPE-LQFAVD----YH 599

Qy     679 EDVYVGYRYYEFADKDVNFPFGHGLSYTTFAFSNLSVSHKDGKLSVSLSVKNTGSVPGAQ 738
Db     600 EGAAVGYKWFDLKGHKPLFPFGHGLSYTTFAYSGLSGQLKDGRLHVRFKVTNTGNVAGKD 659

Qy     739 VAQLYVKPLQAAKINRPVKELKGFAKVELQPGETKAVTIEEQEKYVA----AYFDEERDQ 794
Db     660 VPQVYAAPM-STKWEAP-KRLAAWSKVALLPGETKPGATGTAVLKIAPRDLAYFDVEAGR 717

Qy     795 WCVEKGDYEVIVSDSS 810
Db     718 FRADAGKYELIVAASA 733


RESULT 10
AAW69761
ID   AAW69761 standard; Protein; 735 AA.
XX
AC   AAW69761;
XX
DT   23-NOV-1998  (first entry)
XX
DE   Acetobacter xylinum beta-glucosidase.
XX
KW   Acetobacter xylinum; sucrofermentans; cellulose synthesis complex;
KW   bcsA; bcsB; bcsC; bcsD; CMCase; beta-glucosidase; enzyme; cellulose;
KW   microorganism.
XX
OS   Acetobacter xylinum.
XX
PN   WO9839455-A1.
XX
PD   11-SEP-1998.
XX
PF   09-OCT-1997;   97WO-JP03633.
XX
PR   04-MAR-1997;   97JP-0063927.
XX
PA   (BIOP-) BIO-POLYMER RES CO LTD.
XX
PI   Hayashi T,  Tahara N,  Tonouchi N,  Tsuchida T,  Yano H;
PI   Yoshinaga F;
XX
DR   WPI; 1998-495854/42.
DR   N-PSDB; AAV52831.
XX
PT   Gene encoding Acetobacter xylinum cellulose synthetase complex -
PT   containing a group of genes including those for conventional and
PT   novel ^b-glucosidases
XX
PS   Claim 8; Pages 36-37; 50pp; Japanese.
XX
CC   This represents the amino acid sequence of a Acetobacter xylinum
CC   subspecies sucrofermentans beta-glucosidase. The invention provides
CC   a gene encoding a Acetobacter xylinum subspecies sucrofermentans derived
CC   cellulose synthesis complex-produced protein. The gene sequence
CC   represents bcsA, bcsB, bcsC or bcsD, CMCase and a beta-glucosidase
CC   encoding gene. The novel gene and the enzyme participate in the synthesis
CC   of cellulose by microorganisms. Cells transformed with the genes may be
CC   used in the production of cellulose.
XX
SQ   Sequence   735 AA;

  Query Match             19.6%;  Score 860;  DB 19;  Length 735;
  Best Local Similarity   28.3%;  Pred. No. 8.8e-68;
  Matches  241;  Conservative  125;  Mismatches  277;  Indels  210;  Gaps   22;

Qy       2 ADIDVEAILKKLTLAEKVDLLAGID---------------FWHTKALPKHGVPSLRFTD 45
Db      37 ADARARQVLASMSLEDKMSLLFSVDGGGFNGSVAPPGGLGSAAYLRAPQGSGLPDLQISD 96

Qy      46 GPNGVRGTKFF--NGVPAACFPCGTSLGSTFNQTLLEEAGKMMGKEAIAKSAHVILGPTI 103
Db      97 AGLGVRNPAHIRRNG-EAVSLPSGQSTASTWDMDMARQAGVMIGREAWQSGFNILLGGGA 155

Qy     104 NMQRSPLGGRGFESIGEDPFLAGLGAAALIRGIQSTGVQATIKHFLCNDQEDRRMMVQSI 163
Db     156 DLTRDPRGGRNFEYAGEDPLQTGRMVGSTIAGVQSQHVISTLKHYAMNDLETSRMTMSAD 215
```

```
Qy   164 VTERALREIYALPFQIAVRDSQPGAFMTAYNGINGVSCSENPKYLDGMLRKEWGWDGLIM 223
Db   216 IDPVAMRESDLLGFEIALETGHPGAVMCSYNRVNDLYACENPYLLNKTLKQDWHYPGFVM 275

Qy   224 SDWYGTYSTTEAVVAGLDLEMPGP-----PRFRGETLKFNVSNGK-PFIHVIDQRAREVL 277
Db   276 SDWGATHSSARAALAGLDQESAGDHTDARPYFR-TLLAADVKAGRVPEARINDMAER--- 331

Qy   278 QFVKKCAASGVTEN----GPETTVNNTPETAALLRKVGNEGIVLLKNENNVLPLSKKKKT 333
Db   332 -VVRALFAAGLVDHPAQRGPLDVVTDT----LVAQKDEEEGAVLLRNQGNILPLSPTARI 386

Qy   334 LIVGPNAKQATYHGGGSAALRAYYAVTPFDGLSKQLETPPSYTVGAYTTVPPILGEQCLT 393
Db   387 AVIGGHADAGVISGGGS----------------------------SQVDPIGGE---- 412

Qy   394 PDGAPGMRWRVFNEPPGTPNRQHIDELFFTKTDMHLVDYYHPKAADTWYADMEGTYTADE 453
Db   413 ------------------------------------------------------------ 412

Qy   454 DCTYELGLVVCGTAKAYVDDQLVVDNATKQVPGDAFFGSATREETGRINLVKGNTYKFKI 513
Db   413 --------AVKGPGK------------KEWPGDPVY---------------------- 428

Qy   514 EFGSAPTYTLKGDTIVPGHGSLRVGGCKVIDDQAEIEKSVALAKEHDQVIICAGLNADWE 573
Db   429 -FPSSPLKAMQAE--APG---ARI----TYDPGTSIASAVRAARAADVVVVYA---TQFT 475

Qy   574 TEGADRASMKLPGVLDQLIADVAAANPNTVVVMQTGTPEEMPWLDATPAVIQAWYGGNET 633
Db   476 FEGMDAPSMHLDDNADALITAVAAANPRTVVVMETGDPVLMPWNSSVAGVLEAWFPGSGG 535

Qy   634 GNSIADVVFGDYNPSGKLSLSFPKR------------LQDNPAFLNFRTEAGRTLYGEDV 681
Db   536 GPAIARLLFGKVAPSGHLTMTFPQAESQLAHPDIAGVTADNVFEMQFHTDQ-ELVYDEGS 594

Qy   682 YVGYRYYEFADKDVNFPFGHGLSYTTFAFSNLSVSHKDGKLSVSLSVKNTGSVPGAQVAQ 741
Db   595 DVGYRWFDRNHFKPLYPFGYGLTYTTFSTDGLKVTERHGQVTATFNVHNTGTRAGVDVPQ 654

Qy   742 LYVKPLQAAKINRPVKELKGFAKVELQPGETKAVTIEEQEKYVAAYFDEERDQWCVEKGD 801
Db   655 VYV-----GLPDGGARRLAGWQRISLAPGESRQVSV-QLEPRLLAHFDGKHDRWSVPSGT 708

Qy   802 YEVIVSDSSAAKD 814
Db   709 FRVWL--ASCATD 719


RESULT 11
AAR85197
ID   AAR85197 standard; Protein; 803 AA.
XX
AC   AAR85197;
XX
DT   25-JUN-1996  (first entry)
XX
DE   Tomatinase - a saponin glycosyl hydrolase.
XX
KW   saponin glycosyl hydrolase; tomatinase; plant pathogenic fungi;
KW   avenacinase; deglycosylation; pore formation; cell death.
XX
OS   Septoria lycopersici.
XX
PN   WO9530009-A2.
XX
PD   09-NOV-1995.
XX
PF   17-MAR-1995;   95WO-GB00592.
XX
PR   29-APR-1994;   94GB-0008573.
XX
PA   (GATS-) GATSBY CHARITABLE FOUND.
XX
PI   Bowyer P,  Daniels MJ,  Osbourn AE;
XX
DR   WPI; 1995-393080/50.
DR   N-PSDB; AAT06021.
XX
PT   New isolated saponin glycosyl hydrolase enzymes - used to develop
PT   prods. for the modification of microbial organisms and plants or
PT   plant prods.
XX
PS   Claim 2; Fig 4b; 113pp; English.
XX
CC   AAR85197 is tomatinase, a saponin glycosyl hydrolase (SGH) encoded by
CC   AAT06021, isolated from the plant pathogenic fungi Septoria lycopersici.
CC   The action of tomatinase is mechanistically similar to that of
CC   avenacinase (AAR85198), and involves hydrolytic cleavage of the terminal
CC   beta, 1-2 linked glucose from alpha-tomatine.  This deglycosylation is
CC   sufficient to destroy the ability of the saponin to complex with
CC   membrane sterols. Saponin/sterol complexes in eukaryotic membranes
CC   results in pore formation and leakage of cell contents, with subsequent
CC   cell death. The DNA and proteins of the invention are useful in
CC   identification of related enzymes, structural studies of saponins and
CC   also for development of agents which can modulate SGH activity, e.g. for
CC   reducing pathogenicity of SGH-producing pathogens for specific hosts.
XX
SQ   Sequence   803 AA;

  Query Match             18.1%;  Score 796.5;  DB 16;  Length 803;
  Best Local Similarity   28.3%;  Pred. No. 5.3e-62;
  Matches  253;  Conservative  109;  Mismatches  283;  Indels  249;  Gaps   28;

Qy    10 LKKLTLAEKVDLLAGIDFWHTKALPKHG----VPSLRFT-----DGPNGVRGTKFFNGVP 60
Db    54 VSRLNLTQKVALTTGT----TAGLSCNGNIAPIPEINFSGLCLADGPVSVRIADL----- 104

Qy    61 AACFPCGTSLGSTFNQTLLEEAGKMMGKEAIAKSAHVILGPTIN-MQRSPLGGRGFESIG 119
Db   105 ATVFPAGLTAAATWDRQLIYERARALGSEFRGKGSQVHLGPASGALGRHPLGGRNWESFS 164

Qy   120 EDPFLAGLGAAALIRGIQSTGVQATIKHFLCNDQEDRR----------MNVQSIVTERA 168
Db   165 PDPYLSGVAMDFSIRGIQEMGVQANRKHFIGNEQETQRSNTFTDDGTEIQAISSNIDDRT 224

Qy   169 LREIYALPFQIAVRDSQPGAFMTAYNGINGVSCSENPKYLDGMLRKEWGWDGLIMSDWYG 228
Db   225 MHELYLWPFANAVR-SGVASVMCSYNRLNQTYACENSKLMNGILKGELGFQGYVVSDWYA 283

Qy   229 TYSTTEAVVAGLDLEMPG----------PPRFRGETLKFNVSNGKPFIHVIDQRAREVL 277
Db   284 THSGVESVNAGLDMTMPGPLDSPSTALRPPPSYLGGNLTEAVLNGTIPEARVDDMARRIL 343

Qy   278 ----------QFVKKCAASGV----TENGPETTVN-------NTPET-------AALLRK 309
Db   344 MPYFFLGQDTDFPTVDPSTGFVFARTYNYPDEYLTLGGLDPYNPPPARDVRGNHSDIVRK 403

Qy   310 VGNEGIVLLKNENNVLPLSKKKKTLIVGPNAKQATYHGGGSAALRAYYAVTPFDGLSKQL 369
Db   404 VAAAGTVLLKNVNNVLPLKEPKSVGIFGNGAADVT----------------EGLT--- 442

Qy   370 ETPPSYTVGAYTTVPPILGEQCLTPDGAPGMRWRVFNEPPGTPNRQHIDELFFTKTDMHL 429
Db   443 ------------------------------------------------FTGDD--- 447

Qy   430 VDYYHPKAADTWYADMEGTYTADEDCTYELGLVVCGTAKAYVDDQLVVDNATKQVPGDAF 489
Db   448 --------SGPWGADIG------------ALSVGGGSGAGRHTHLVSPLA--------- 477

Qy   490 FGSATREETGRINLVKGNTYKFKIEFGSAPTYTLKGDTIVPGHGSLRVGGCKVIDDQAEI 549
Db   478 ---AIRKRTESV-------------GGRVQYLLSNSRIVN------------DDFTSI 507

Qy   550 EKSVALAKEHDQVIICAGLNADWETEGADRASMKLPGVLDQLIADVAAANPNTVVVMQTG 609
```

```
                :    :|      |  || || | :       :: :||   |||:||  :|
Db    508 ---------YPTPEVCLVFLKTWAREGTDRLSYENDWNSTAVVNNVARRCPNTIVVTHSG 558

Qy    610 TPEEMPWLD--ATPAVIQAWYGGNETGNSIADVVFGDYNPSGKLSLSFPKRLQD-NPAFL 666
          ||| |     |:: | | | | |||| |:::|| ||||:|  : ||   | :   :
Db    559 GINTMPWADNANVTAILAAHYPGQENGNSIMDILYGDVNPSGRLPYTIPKLATDYDFPVV 618

Qy    667 NFRTEA-----GRTLYGEDVYVGYRYYEFADKDVNFPFGHGLSYTTF------------ 708
          |   ||      :  : | : : ||:::  :    : ||:|||||||
Db    619 NITNEAQDPYVWQADFTEGLLIDYRHFDARNITPLYEFGYGLSYTTFEIEGVANLVAKSA 678

Qy    709 ---AF-SNLSVSHKDGK-------LSVSLSVKNTGSVPGAQVAQLYVK-PLQAAKINRPV 756
             || ::  :||  |        :||: :||||||| |:|| |||:  |      | |:
Db    679 KLSAFPASTDISHPGGNPDLWEEVVSVTAAVKNTGSVSGSQVVQLYISLPADGIPENSPM 738

Qy    757 KELKGFAKVELQPGETKAVTIEEQEKYVAAYFDEERDQWCVEKGDYEVIVSDSS 810
          : |:|| ||:|||||::|:|     : : ::::     | :  | |  |  ||
Db    739 QVLRGFEKVDLQPGQSKSVEFSIMRRDL-SFWNTTAQDWEIPNGQIEFRVGFSS 791


RESULT 12
AAW35004
ID   AAW35004 standard; Protein; 778 AA.
XX
AC   AAW35004;
XX
DT   21-MAY-1998  (first entry)
XX
DE   Thermotoga maritima endoglucanase.
XX
KW   Endoglucanase; cellulase; carboxymethylcellulose; cellulose;
KW   biomass; beta-1,4-glycosidic bond; hydrolysis; saccharification;
KW   thermostable enzyme; thermophilic; glycosidase.
XX
OS   Thermotoga maritima strain MSB8 (Clone 6GC18).
XX
PN   WO9744361-A1.
XX
PD   27-NOV-1997.
XX
PF   22-MAY-1997;   97WO-US08793.
XX
PR   22-MAY-1996;   96US-0651572.
XX
PA   (RECO-) RECOMBINANT BIOCATALYSIS INC.
XX
PI   Lam DE,  Mathur EJ;
XX
DR   WPI; 1998-018435/02.
DR   N-PSDB; AAT94212.
XX
PT   Endoglucanase(s), preferably form archael bacterium, AEPII 1a -
PT   useful to degrade carboxymethylcellulose and hydrolyse of
PT   beta-1,4-glycosidic bonds in cellulose
XX
PS   Claim 1; Fig 1T; 164pp; English.
XX
CC   This protein comprises an endoglucanase of Thermotoga maritima MSB8
CC   (Clone 6GC18) that is capable of degrading carboxymethylcellulose
CC   and of hydrolysing the beta-1,4-glycosidic bonds in cellulose.  It
CC   has homology to an endoglucanase of archaebacterium AEPII1a (see
CC   AAW34985).  It can be produced from native cells or from recombinant
CC   host cells, especially prokaryotic host cells transformed with a
CC   plasmid or virus-derived vector including the endoglucanase DNA
CC   (see AAT94212).  24 Endoglucanases (see AAW34986-W35008) are claimed.
CC   They can be used to degrade cellulose for the conversion of plant
CC   biomass into fuels and chemicals, for use in detergents, textiles,
CC   animal feed, waste treatment, and in the fruit juice and brewing
CC   industries for the clarification and extraction of juices.
XX
SQ   Sequence   778 AA;

  Query Match             17.0%;  Score 747;  DB 19;  Length 778;
  Best Local Similarity   26.1%;  Pred. No. 1.4e-57;
  Matches  236;  Conservative  149;  Mismatches  285;  Indels  234;  Gaps   34;

Qy      3 DIDVEAILKKLTLAEKVDLLAGIDFW-------------HTKALPKHGVPSLRFTDGPN 48
          :: |  :| ::|| ||| |  : |                 | | |:|:  :    |
Db     12 EVRVRDLLSRMTLEEKVAQLGSV--WGYELIDERGKFSREKAKELLKNGIGQITRPGGST 69

Qy     49 GV---RGTKFFN------------GVPA----AC-----------FPCGTSLGSTFNQTL 78
          :      : |             |:||     |           ||   :: ||::  |
Db     70 NLEPQEAAELVNEIQRFLVEETRLGIPAMIHEECLTGYMGLGGTNFPQAIAMASTWDPDL 129

Qy     79 LEEAGKMMGKEAIAKSAHVILGPTINMQRSPLGGRGFESIGEDPFLAGLGAAALIRGIQ- 137
          :|:    : ::     ||  | | ::: | |  ||  |: || |:|      : ::|:|
Db    130 IEKMTTAVREDMRKIGAHQGLAPVLDVARDPRWGRTEETFGESPYLVARMGVSYVKGLQG 189

Qy    138 ---STGVQATIKHFL-CNDQEDRRMMVQSIVTERALREIYALPFQIAVRDSQPGAFMTAY 193
             || ||:|||   :  |  :    : : ||  :|::  ||: ||:::   : | :|
Db    190 EDIKKGVVATVKHFAGYSASEGGKNWAPTNIPEREFKEVFLFPFEAAVKEANVLSVMNSY 249

Qy    194 NGINGVSCSENPKYLDGMLRKEWGWDGLIMSDWYG-------------TYSTTEAVVAG 239
          : |:|| |: | | |  :|||:||::|:::||::                : :   |: ||
Db    250 SEIDGVPCAANRKLLTDILRKDWGFEGIVVSDYFAVKVLEDYHRIARDKSEAARLALEAG 309

Qy    240 LDLEMPGPPRFRGETLKFNVSNGKPFIHVIDQRAREVLQFVKKCAASGVTENGPETTVNN 299
          :|:|:|    :  : ||  | |     :||:    ||:         |: || |   |
Db    310 IDVELPKTECY--QYLKDLVEKGIISEALIDEAVTRVLRLK---FMLGLFEN-PYVEVEK 363

Qy    300 T--PETAALLRKVGNEGIVLLKNENNVLPLSKKKKTLIVGPNAKQATYHGGGSAALRAYY 357
                :  ::  : |:|||:  :||| | ||  ::|||| :
Db    364 AKIESHRDIALEIARKSIILLKND-GILPLQKNKKVALIGPNAGE--------------- 407

Qy    358 AVTPFDGLSKQLETPPSYTVGAYTTVPPILGEQCLTPDGAPGMRWRVFNEPPGTPNRQHI 417
                                  |  :||:                          ||
Db    408 -------------------------VRNLLGDYMYL---------------------AHI 421

Qy    418 DELFFTKTDMHLVDYYHPKAADTWYADMEGTYTADEDCTYELGLVVCGTAKAYVDDQLVV 477
            |     |:    : :|:     |  :: :                           :
Db    422 RALLDNIDDV----FGNPQIPRENYERLKKS---------------------------I 449

Qy    478 DNATKQVPG--DAFFGSATREETGRINLVKGNTYKFKIEFGSAPTYTLKGDTIVPGHGSL 535
          :   | :|   |||      :| |              |||  |
Db    450 EEHMKSIPSVLDAF------KEEG-------------IEFEYAK---------------- 474

Qy    536 RVGGCKVI-DDQAEIEKSVALAKEHDQVIIC----AGLNADWET-EGADRASMKLPGVLD 589
             ||:|  :|::  |::: :||: |  |:      :|  |  | |  | |::|||| :
Db    475 ---GCEVTGEDRSGFEEAIEIAKKSDVAIVVVGDKSGLTLDCTTGESRDMANLKLPGVQE 531

Qy    590 QLIADVAAANPNTVVVMQTGTPEEMP-WLDATPAVIQAWYGGNETGNSIADVVFGDYNPS 648
          :|: :||      |:|: || |  :   :|   |::| |  |   | :| |:::|  |||
Db    532 ELVLEVAKTGKPVVLVLITGRPYSLKNVVDKVNAILQVWLPGEAGGRAIVDIIYGKVNPS 591

Qy    649 GKLSLSFPKRLQDNPAFLNFRTEAGRTLYGEDVYVGYRYYEFADKDVNFPFGHGLSYTTF 708
          ||| :|||:     | |   :   ||: :  |       | : : | : ||||||||| |
Db    592 GKLPISFPRSAGQIPVFHYVKPSGGRSHWHGD------YVDESTKPL-FPFGHGLSYTKF 644

Qy    709 AFSNLSVSHKD----GKLSVSLSVKNTGSVPGAQVAQLYVKPLQAAKINRPVKELKGFAK 764
          :|||  :  |:    |:: : : |:| |   | :| |||:   : | : |||||||| :
Db    645 EYSNLRIEPKEVPPAGEVVIKVDVENIGDRDGDEVVQLYI-GREFASVTRPVKELKGFKR 703

Qy    765 VELQPGETKAVTIEEQEKYVAAYFDEERD-QWCVEKGDYEVIVSDSSAAKDGVALRGKFT 823
          | |:  | | |         :| |    || :  || |:::|:|  ||    : :  | | |:
Db    704 VSLKAKEKKTVVFRLHMDVLAYY---NRDMKLVVEPGEFKVMVGSSS---EDIRLTGSFS 757

Qy    824 -VGE 826
           |||
Db    758 VVGE 761


RESULT 13
AAR25384
ID   AAR25384 standard; Protein; 744 AA.
XX
AC   AAR25384;
```

```
XX
DT   06-JAN-1993  (first entry)
XX
DE   Trichoderma reesei B-gal.
XX
KW   Beta-galactosidase; filamentous fungi; cellulase; detergent;
KW   cellooligosaccharide; grapes; wine; feedstock; biomass; bgl1.
XX
OS   Trichoderma reesei.
XX
PN   WO9210581-A.
XX
PD   25-JUN-1992.
XX
PF   10-DEC-1991;   91WO-US09285.
XX
PR   10-DEC-1990;   90US-0625140.
XX
PA   (GEMV ) GENENCOR INT INC.
XX
PI   Barnett CC,  Fowler T,  Shoemaker S;
XX
DR   WPI; 1992-234636/28.
DR   P-PSDB; AAR25384.
XX
PT   Extracellular beta-glucosidase expression in filamentous fungi -
PT   enhances cellulose degradation in feedstock, biomass and sludge
XX
PS   Disclosure; Fig 1; 101pp; English.
XX
CC   The beta-galactosidase amino acid sequence was deduced from the DNA
CC   sequence of the bgl1 gene obtd. using PCR primers based on the N-
CC   terminal region of bgl1 gene for amplifying a cDNA library to recover
CC   a 700 bp bgl1 clone which was labelled and used to probe a T.
CC   reesei genomic DNA library to recover the full length bgl1 gene.
CC   Bgl1 encodes the 74.3 kD protein beta galactosidase. Transformants
CC   of T. reesei can be used to produce fungal cellulase compsns. B-gal
CC   can be isolated from the culture medium of enriched transformants
CC   and added to grapes during wine making, to enhance the potential
CC   aroma of the finished wine prod. B-gal may also be used in fruit
CC   to enhance the aroma.  Enhanced B-gal may be used to degrade
CC   cellulosics in feedstock, biomass and sludge.  The transformants
CC   may be used in detergent compsns. to isolate cellooligosaccharides
CC   and may be useful as food additives, chemical intermediates, etc.
XX
SQ   Sequence   744 AA;

  Query Match             16.7%;  Score 731.5;  DB 13;  Length 744;
  Best Local Similarity   27.5%;  Pred. No. 3.2e-56;
  Matches  234;  Conservative  123;  Mismatches  282;  Indels  211;  Gaps   31;

Qy       7 EAILKKLTLAEKVDLLAGIDFW-------HTKALPKHGVPSLRFTDGPNGVR----GTKF 55
           :| | || | :|| :::|:  |        :|    |  |||   ||| |||     | |
Db      48 KAALAKLNLQDKVGIVSGVG-WNGGPCVGNTSPASKISYPSLCLQDGPLGVRYSTGSTAF 106

Qy      56 FNGVPAACFPCGTSLGSTFNQTLLEEAGKMMGKEAIAKSAHVILGPTIN-MQRSPLGGRG 114
             || ||          ||::  |: | |: :|:|  |   ||||||    : ::| |||
Db     107 TPGVQAA---------STWDVNLIRERGQFIGEEVKASGIHVILGPVAGPLGKTPQGGRN 157

Qy     115 FESIGEDPFLAGLGAAALIRGIQSTGVQATIKHFLCNDQEDRRMMVQSIVTERALREIYA 174
           :|  | ||:| |:     | |||| ||||| ||:: |:|| |  : |   :| | |:|
Db     158 WEGFGVDPYLTGIAMGQTINGIQSVGVQATAKHYILNEQELNRETISSNPDDRTLHELYT 217

Qy     175 LPFQIAVRDSQPGAFMTAYNGINGVSCSENPKYLDGMLRKEWGWDGLIMSDWYGTYSTTE 234
             ||  ||: :   : | :|| :|     |:   |  :|: : |: | :|:||   ::| :
Db     218 WPFADAVQ-ANVASVMCSYNKVNTTWACEDQYTLQTVLKDQLGFPGYVMTDWNAQHTTVQ 276

Qy     235 AVVAGLDLEMPGPP-----RFRGETLKFNV-SNGKPFIHVIDQRAREVLQFVKKCAASGV 288
           :  :|||: |||       |  | |   | || |   | |     | : :     :|
Db     277 SANSGLDMSMPGTDFNGNNRLWGPALTNAVNSNQVPTSRVDDM----VTRILAAWYLTGQ 332

Qy     289 TENG-PETTVNNTPE--TAALLRKVGNEGIVLLKNENNVLPLSKKKKTLIVGPNAKQATY 345
           : | |   ::   :      :| :  :|||||||: |:||| |     :|

Db     333 DQAGYPSFNISRNVQGNHKTNVRAIARDGIVLLKNDANILPLKKPASIAVV--------- 383

Qy     346 HGGGSAALRAYYAVTPFDGLSKQLETPPSYTVGAYTTVPPILGEQ-CLTPDGAPGMRWRV 404
             ||||:                        :| :    |   :: |   ||| ;| |
Db     384 ---GSAAI----------------------IGNHARNSPSCNDKGC--DDGALGMGW-- 413

Qy     405 FNEPPGTPNRQHIDELFFTKTDMHLVDYYHPKAADTWYADMEGTYTADEDCTYELGLVVC 464
             | |  :               :  | |    |  :||
Db     414 ---GSGAVNYPY--------------FVAPYDAINTRASSQGT---------------- 439

Qy     465 GTAKAYVDDQLVVDNATKQVPGDAFFGSATREETGRINLVKGNTYKFKIEFGSAPTYTLK 524
                 |: : |      |     || | :   |  :  ::
Db     440 ---------QVTLSNTDNTSSG----ASAARGKDVAIVFITADS---------------- 470

Qy     525 GDTIVPGHGSLRVGGCKVIDDQAEIEKSVALAKEHDQVIICAGLNADWETEGADRASMKL 584
               | : | |      |:  ::                    |    |
Db     471 ------GEGYITVEG--NAGDRNNLD-------------------PWHNGNA------- 495

Qy     585 PGVLDQLIADVAAANPNTVVVMQTGTP---EEMPWLDATPAVIQAWYGGNETGNSIADVV 641
               |:  || || | :||: :      |::  |    ||: |     |:||:: ||:
Db     496 ------LVQAVAGANSNVIVVVHSVGAIILEQILALPQVKAVVWAGLPSQESGNALVDVL 549

Qy     642 FGDYNPSGKLSLSFPKRLQDNPAFLNFR-TEAGRTLYGEDVYVGYRYYEFADKDVNFPFG 700
           :|| :|||||  :  |    :|   | |    |   : | ::: |:::: |:    : ||
Db     550 WGDVSPSGKLVYTIAK----SPNDYNTRIVSGGSDSFSEGLFIDYKHFDDANITPRYEFG 605

Qy     701 HGLSYTTFAFSNLSV--SHKDGK----------------LSVSLSVKNTGSVPGAQVAQ 741
           :|||||  | :|  |||   : | |                :|:: : |:|  | ||:|||
Db     606 YGLSYTKFNYSRLSVLSTAKSGPATGAVVPGGPSDLFQNVATVTVDIANSGQVTGAEVAQ 665

Qy     742 LYVK-PLQAAKINRPVKELKGFAKVELQPGETKAVTIEEQEKYVAAYFDEERDQWCVEKG 800
           ||:  |   | :   | |:|:|||: | ||::   |   : : : :|:|    :| |  |
Db     666 LYITYPSSAPR--TPPKQLRGFAKLNLTPGQSGTATFNIRRRDL-SYWDTASQKWVVPSG 722

Qy     801 DYEVIVSDSS 810
             : : |  ||
Db     723 SFGISVGASS 732


RESULT 14
AAB08340
ID   AAB08340 standard; Protein; 744 AA.
XX
AC   AAB08340;
XX
DT   04-DEC-2000  (first entry)
XX
DE   Amino acid sequence of a beta-glucosidase polypeptide.
XX
KW   Beta-glucosidase; bg11 gene; filamentous fungus.
XX
OS   Trichoderma reesei.
XX
PN   US6103464-A.
XX
PD   15-AUG-2000.
XX
PF   05-JUN-1995;   95US-0463461.
XX
PR   24-MAY-1994;   94US-0248586.
PR   10-DEC-1991;   91US-0807028.
PR   10-DEC-1990;   90US-0625140.
XX
PA   (GEMV ) GENENCOR INT INC.
XX
PI   Shoemaker S,  Barnett CC,  Fowler T;
XX
DR   WPI; 2000-557671/51.
DR   N-PSDB; AAA63953.
XX
PT   Detecting DNA encoding beta-glucosidase from filamentous fungus
PT   comprises hybridizing fungal DNA with a nucleotide sequence encoding
PT   Trichoderma reesei beta-glucosidase and detecting the hybridized DNA  -
```

```
XX
PS   Disclosure; Fig 1; 36pp; English.
XX
CC   The present sequence represents a beta-glucosidase. The beta-glucosidase
CC   polypeptide is extracellular or cell wall bound protein encoded by the
CC   bgl1 gene. The specification describes a method for detecting DNA
CC   encoding beta-glucosidase from a filamentous fungus. The method
CC   comprises contacting the DNA of the fungi with a Trichoderma reesei
CC   bgl1 polynucleotide, and detecting hybridised DNA sequences. The method
CC   is used for detecting DNA encoding beta-glucosidase from filamentous
CC   fungi. Once homologous genes of other filamentous fungi are cloned,
CC   they can be used to transform the filamentous fungi to overproduce
CC   beta-glucosidase. Alternatively, the cloned bgl1 genes can be used to
CC   delete or disrupt the bgl1 genes.
XX
SQ   Sequence   744 AA;

  Query Match             16.7%;  Score 731.5;  DB 21;  Length 744;
  Best Local Similarity   27.5%;  Pred. No. 3.2e-56;
  Matches  234;  Conservative  123;  Mismatches  282;  Indels  211;  Gaps   31;

Qy       7 EAILKKLTLAEKVDLLAGIDFW-------HTKALPKHGVPSLRFTDGPNGVR----GTKF 55
           :| | || | :|| :::|:  |        :|    |   |||   ||| |||     | |
Db      48 KAALAKLNLQDKVGIVSGVG-WNGGPCVGNTSPASKISYPSLCLQDGPLGVRYSTGSTAF 106

Qy      56 FNGVPAACFPCGTSLGSTFNQTLLEEAGKMMGKEAIAKSAHVILGPTIN-MQRSPLGGRG 114
             || ||         ||::  |: | |: :|:|  |   ||||||    : ::| |||
Db     107 TPGVQAA---------STWDVNLIRERGQFIGEEVKASGIHVILGPVAGPLGKTPQGGRN 157

Qy     115 FESIGEDPFLAGLGAAALIRGIQSTGVQATIKHFLCNDQEDRRMMVQSIVTERALREIYA 174
           :|  | ||:| |:     | |||| ||||| ||:: |:||  |  : |   :| | |:|
Db     158 WEGFGVDPYLTGIAMGQTINGIQSVGVQATAKHYILNEQELNRETISSNPDDRTLHELYT 217

Qy     175 LPFQIAVRDSQPGAFMTAYNGINGVSCSENPKYLDGMLRKEWGWDGLIMSDWYGTYSTTE 234
             ||  ||: :   : | :|| :|     |:   |  :|: : |: | :|:||   ::| :
Db     218 WPFADAVQ-ANVASVMCSYNKVNTTWACEDQYTLQTVLKDQLGFPGYVMTDWNAQHTTVQ 276

Qy     235 AVVAGLDLEMPGPP-----RFRGETLKFNV-SNGKPFIHVIDQRAREVLQFVKKCAASGV 288
           :  :|||: |||       |  |  |   | ||  |   | |-    | : :   :|
Db     277 SANSGLDMSMPGTDFNGNNRLWGPALTNAVNSNQVPTSRVDDM----VTRILAAWYLTGQ 332

Qy     289 TENG-PETTVNNTPE--TAALLRKVGNEGIVLLKNENNVLPLSKKKKTLIVGPNAKQATY 345
           : | |   ::   :      :| :  :|||||||: |:||| |     :|
Db     333 DQAGYPSFNISRNVQGNHKTNVRAIARDGIVLLKNDANILPLKKPASIAVV--------- 383

Qy     346 HGGGSAALRAYYAVTPFDGLSKQLETPPSYTVGAYTTVPPILGEQ-CLTPDGAPGMRWRV 404
             ||||:                        :| :    |   :: |  ||| || |
Db     384 ---GSAAI----------------------IGNHARNSPSCNDKGC--DDGALGMGW-- 413

Qy     405 FNEPPGTPNRQHIDELFFTKTDMHLVDYYHPKAADTWYADMEGTYTADEDCTYELGLVVC 464
           |  |  :                : :|  |    |  :||
Db     414 ---GSGAVNYPY--------------FVAPYDAINTRASSQGT---------------- 439

Qy     465 GTAKAYVDDQLVVDNATKQVPGDAFFGSATREETGRINLVKGNTYKFKIEFGSAPTYTLK 524
                 |: : |     |     || | :   |  : ::
Db     440 ---------QVTLSNTDNTSSG----ASAARGKDVAIVFITADS---------------- 470

Qy     525 GDTIVPGHGSLRVGGCKVIDDQAEIEKSVALAKEHDQVIICAGLNADWETEGADRASMKL 584
           |  | : | |     |: ::                    |    |
Db     471 ------GEGYITVEG--NAGDRNNLD-------------------PWHNGNA------- 495

Qy     585 PGVLDQLIADVAAANPNTVVVMQTGTP---EEMPWLDATPAVIQAWYGGNETGNSIADVV 641
                 |:  || || | :||: :      |::  |    ||: |     |:||:: ||:
Db     496 ------LVQAVAGANSNVIVVVHSVGAIILEQILALPQVKAVVWAGLPSQESGNALVDVL 549

Qy     642 FGDYNPSGKLSLSFPKRLQDNPAFLNFR-TEAGRTLYGEDVYVGYRYYEFADKDVNFPFG 700
           :|| :|||||  :  |   :|   | |     |   : | ::: |:::: |:    : ||
Db     550 WGDVSPSGKLVYTIAK----SPNDYNTRIVSGGSDSFSEGLFIDYKHFDDANITPRYEFG 605

Qy     701 HGLSYTTFAFSNLSV--SHKDGK-----------------LSVSLSVKNTGSVPGAQVAQ 741
           :|||||  | :| |||  : | |                  :|:: : |:| | |||:||
Db     606 YGLSYTKFNYSRLSVLSTAKSGPATGAVVPGGPSDLFQNVATVTVDIANSGQVTGAEVAQ 665

Qy     742 LYVK-PLQAAKINRPVKELKGFAKVELQPGETKAVTIEEQEKYVAAYFDEERDQWCVEKG 800
           ||:  |  | :   | |:|:|||||: | ||::   |   : : : :|:|    :| |  |
Db     666 LYITYPSSAPR--TPPKQLRGFAKLNLTPGQSGTATFNIRRRDL-SYWDTASQKWVVPSG 722

Qy     801 DYEVIVSDSS 810
             : :  |  ||
Db     723 SFGISVGASS 732


RESULT 15
AAY56548
ID   AAY56548 standard; Protein; 744 AA.
XX
AC   AAY56548;
XX
DT   21-FEB-2000  (first entry)
XX
DE   Trichoderma reesei beta-glucosidase.
XX
KW   Trichoderma reesei; beta-glucosidase; bgl1; filamentous fungus;
KW   cellulase; food; flavour; aroma; wine; fruit.
XX
OS   Trichoderma reesei.
XX
PN   US5997913-A.
XX
PD   07-DEC-1999.
XX
PF   05-JUN-1995;   95US-0462080.
XX
PR   24-MAY-1994;   94US-0248586.
PR   10-DEC-1991;   91US-0807028.
PR   10-DEC-1990;   90US-0625140.
XX
PA   (GEMV ) GENENCOR INT INC.
XX
PI   Shoemaker S,  Barnett CC,  Fowler T;
XX
DR   WPI; 2000-052536/04.
DR   N-PSDB; AAZ38879.
XX
PT   Enhancing the flavor of foods using an enzyme overexpressed by a
PT   recombinant filamentous fungi -
XX
PS   Claim 4; Fig 1; 45pp; English.
XX
CC   A method has been developed of enhancing the flavour of foods using
CC   beta-glucosidase overexpressed by a recombinant filamentous fungi
CC   selected from the genus Trichoderma, Aspergillus, Neurospora, Humicola
CC   and Penicillium. The method comprises: (a) expressing a fungal DNA
CC   sequence encoding beta-glucosidase in a transformed host using an
CC   expression vector, where the DNA sequence is capable of being amplified
CC   by polymerase chain reaction (PCR) with primers GCNGTNCCTCCTGCNGG
CC   (AAZ38880) and GTTRTTNCCRTTRAARTCTGT (AAZ38881) under amplification
CC   conditions of: denaturation at 95 plus degrees Celsius for 10 minutes,
CC   annealing at 50 plus degrees Celsius for 2 minutes and extension at
CC   65 plus degrees Celsius for 10 minutes, for 30 cycles; (b) culturing
CC   the transformants under conditions to permit growth; (c) isolating
CC   beta-glucosidase produced from the transformants; and (d) adding the
CC   beta-glucosidase to foods. The method is useful for enhancing the
CC   flavour or aroma of wine or fruit. The present sequence represents
CC   beta-glucosidase isolated from Trichoderma reesei.
XX
SQ   Sequence   744 AA;

  Query Match             16.7%;  Score 731.5;  DB 21;  Length 744;
  Best Local Similarity   27.5%;  Pred. No. 3.2e-56;
  Matches  234;  Conservative  123;  Mismatches  282;  Indels  211;  Gaps   31;

Qy       7 EAILKKLTLAEKVDLLAGIDFW-------HTKALPKHGVPSLRFTDGPNGVR----GTKF 55
           :| | || | :|| :::|:  |        :|    |   |||   ||| |||     | |
Db      48 KAALAKLNLQDKVGIVSGVG-WNGGPCVGNTSPASKISYPSLCLQDGPLGVRYSTGSTAF 106
```

```
Qy      56 FNGVPAACFPCGTSLGSTFNQTLLEEAGKMMGKEAIAKSAHVILGPTIN-MQRSPLGGRG 114
Db     107 TPGVQAA---------STWDVNLIRERGQFIGEEVKASGIHVILGPVAGPLGKTPQGGRN 157

Qy     115 FESIGEDPFLAGLGAAALIRGIQSTGVQATIKHFLCNDQEDRRMMVQSIVTERALREIYA 174
Db     158 WEGFGVDPYLTGIAMGQTINGIQSVGVQATAKHYILNEQELNRETISSNPDDRTLHELYT 217

Qy     175 LPFQIAVRDSQPGAFMTAYNGINGVSCSENPKYLDGMLRKEWGWDGLIMSDWYGTYSTTE 234
Db     218 WPFADAVQ-ANVASVMCSYNKVNTTWACEDQYTLQTVLKDQLGFPGYVMTDWNAQHTTVQ 276

Qy     235 AVVAGLDLEMPGPP-----RFRGETLKFNV-SNGKPFIHVIDQRAREVLQFVKKCAASGV 288
Db     277 SANSGLDMSMPGTDFNGNNRLWGPALTNAVNSNQVPTSRVDDM----VTRILAAWYLTGQ 332

Qy     289 TENG-PETTVNNTPE--TAALLRKVGNEGIVLLKNENNVLPLSKKKKTLIVGPNAKQATY 345
Db     333 DQAGYPSFNISRNVQGNHKTNVRAIARDGIVLLKNDANILPLKKPASIAVV--------- 383

Qy     346 HGGGSAALRAYYAVTPFDGLSKQLETPPSYTVGAYTTVPPILGEQ-CLTPDGAPGMRWRV 404
Db     384 ---GSAAI----------------------IGNHARNSPSCNDKGC--DDGALGMGW-- 413

Qy     405 FNEPPGTPNRQHIDELFFTKTDMHLVDYYHPKAADTWYADMEGTYTADEDCTYELGLVVC 464
Db     414 ---GSGAVNYPY--------------FVAPYDAINTRASSQGT---------------- 439

Qy     465 GTAKAYVDDQLVVDNATKQVPGDAFFGSATREETGRINLVKGNTYKFKIEFGSAPTYTLK 524
Db     440 ---------QVTLSNTDNTSSG----ASAARGKDVAIVFITADS---------------- 470

Qy     525 GDTIVPGHGSLRVGGCKVIDDQAEIEKSVALAKEHDQVIICAGLNADWETEGADRASMKL 584
Db     471 ------GEGYITVEG--NAGDRNNLD-------------------PWHNGNA------- 495

Qy     585 PGVLDQLIADVAAANPNTVVVMQTGTP---EEMPWLDATPAVIQAWYGGNETGNSIADVV 641
Db     496 ------LVQAVAGANSNVIVVVHSVGAIILEQILALPQVKAVVWAGLPSQESGNALVDVL 549

Qy     642 FGDYNPSGKLSLSFPKRLQDNPAFLNFR-TEAGRTLYGEDVYVGYRYYEFADKDVNFPFG 700
Db     550 WGDVSPSGKLVYTIAK----SPNDYNTRIVSGGSDSFSEGLFIDYKHFDDANITPRYEFG 605

Qy     701 HGLSYTTFAFSNLSV--SHKDGK----------------LSVSLSVKNTGSVPGAQVAQ 741
Db     606 YGLSYTKFNYSRLSVLSTAKSGPATGAVVPGGPSDLFQNVATVTVDIANSGQVTGAEVAQ 665

Qy     742 LYVK-PLQAAKINRPVKELKGFAKVELQPGETKAVTIEEQEKYVAAYFDEERDQWCVEKG 800
Db     666 LYITYPSSAPR--TPPKQLRGFAKLNLTPGQSGTATFNIRRRDL-SYWDTASQKWVVPSG 722

Qy     801 DYEVIVSDSS 810
Db     723 SFGISVGASS 732
```

Search completed: April 26, 2003, 13:08:21
Job time : 88 secs